Australian Government
Department of Defence
Defence Science and
Technology Organisation

Intelligent Multi-Media Presentation Using Rhetorical Structure Theory

Steven Wark and Marcin Nowina-Krowicki

Joint and Operations Analysis Division
Defence Science and Technology Organisation

DSTO-TR-3067

ABSTRACT

This report explores the feasibility of an Intelligent Multi-Media Presentation (IMMP) system for human-authored content, marked up using Rhetorical Structure Theory, to support dynamic selection of the presentation content based on user needs and preferences. It describes an XML format developed to represent an IMMP presentation, and a simplified prototype system developed to dynamically select and present content at different levels of detail within a specified maximum duration. An initial assessment of this system, based on the TTCP 'Military Strikes in Atlantis' scenario, found that it performed satisfactorily and that this is thus a feasible approach. Further work is planned to assess this system with other scenarios, and determine if it is a generally suitable approach.

RELEASE LIMITATION

Approved for public release

Published by

Joint and Operations Analysis Division
DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend, Victoria 3207 Australia

Telephone: 1300 333 362
Fax: (03) 9626 7999

AR-016-212
January 2015

APPROVED FOR PUBLIC RELEASE

Intelligent Multi-Media Presentation Using Rhetorical Structure Theory

Executive Summary

The era of 'Big Data' has resulted in the exponential growth of sensor and other information within defence and national information systems. As more and more information becomes available, the challenge for users is not only to find relevant information in a timely manner, but to also have it integrated with related information and presented to them in a contextually appropriate form. The goal of Intelligent Multi-Media Presentation (IMMP) systems is to automatically discover and assemble content in this way.

This report describes initial work towards this goal, looking at the feasibility of an approach based on Rhetorical Structure Theory (RST) to mark up human authored content so that it can be presented in different ways to different audiences based on their current information needs. In particular, this allows the level of detail provided in a presentation to be managed so that its running time is less than a specified maximum duration, while retaining overall narrative coherence. While this is an early step towards the goal of a fully automated IMMP system, it provides an immediately useful capability for re-use of briefing and training materials with different audiences.

In this work, an XML format has been developed to support authoring and mark-up of an IMMP presentation using RST relations, which describe how elements of the presentation narratively relate to other elements. The XML document is structured as one or more sequences of multimedia clips, each of which represents a self-contained set of multimedia information to be presented (akin to a single slide in a slide-pack). Each clip is assembled from one or more multimedia segments, which represent coordinated multimedia builds. The RST relations provide a cue to determine which elements are core to the presentation goals, and which are elaborations that can be dropped without destroying the coherence of the presentation. The XML document is independent of the final presentation mechanism, so that how content is rendered is determined purely by the display devices' capabilities. The RST relations potentially help select which content to retain on less capable devices.

A prototype IMMP system has been developed based on an XML pipeline to select and present IMMP content created in this XML format, using a web-based IMMP authoring tool currently under development. For this early work, a simple selection scheme has been tested that assigns weights to multimedia content based on the RST relations

associated with it. Content is scored in the range [0, 1] by multiplying its weight by the score of any dependencies. An IMMP presentation can then be generated by either filtering out all content with a score less than a nominated threshold, or by finding the selection threshold that produces a presentation with a specified maximum run-time. This quite simple scheme has been found to give promising results for an example IMMP presentation based on the TTCP 'Military Strikes in Atlantis' scenario, using two different weighting schemes: one suitable for a 'naïve' audience that favours retaining the overall presentation structure; and another suitable for a 'expert' audience already familiar with the topic of the presentation, that favours retaining the 'core' (as identified by its RST relation) content over supporting material.

Further work is planned to assess how this system performs with a wider range of scenarios, but this initial work has established the feasibility of this approach. When coupled with DSTO's Virtual Adviser technology for presentation of the material, this system has the potential to allow routine briefing and training materials to be prepared only once, and then to be re-used many times, tuned for different audiences and time constraints.

One of the assumptions made with this work is that the IMMP presentation has been created and marked-up by a human author. It is anticipated that a fully automated system that generates the IMMP presentation from source material can assign RST relations at creation time, and use a similar scoring system to select the presentation material appropriate for different audiences and time constraints.

Authors

Dr Steven Wark
Joint and Operations Analysis Division

Steve Wark was awarded a Ph.D in physics from the University of Melbourne in 1992, and joined the Microwave Radar Division of DSTO in 1996 as a Research Scientist. In 2000 he moved to the then Information Technology Division as a Senior Research Scientist, where he worked on the Future Operations Centre Analysis Laboratory (FOCAL) and information fusion systems for command and control. This work involved him in the development of distributed multi-agent systems for information fusion and novel interfaces for C2, such as embodied conversational agents, immersive battlespace displays, and 3D animation.

More recently he has worked on Situational Awareness technologies supporting HQJOC, including R&D towards a Higher Level Common Operating Picture (HiCOP) incorporating aspects of a User Defined Operating Picture (UDOP) and multimedia narrative

Mr Marcin Nowina-Krowicki
Joint and Operations Analysis Division

Marcin Nowina-Krowicki received a bachelor degree in Computer and Information Science from the University of South Australia.

His professional interests lie in novel 2D and 3D user interfaces. His work at DSTO has allowed him to combine his passion for Human Computer Interaction and User Interfaces to help develop embodied conversational agents and multimodal geospatial displays for Situational Awareness for the Future Operations Centre Analysis Laboratory (FOCAL).

More recently he has worked on Situational Awareness technologies supporting HQJOC, including R&D towards a Higher Level Common Operating Picture (HiCOP) incorporating aspects of a User Defined Operating Picture (UDOP) and multimedia narrative.

This page is intentionally blank

Contents

Abbreviations

ADF	Australian Defence Force
API	Application Programming Interface
APMS	Abbreviation and Pronunciation Management System
BIO	Blueland Intelligence Organisation (fictitious)
BIS	Blended Interaction Space(s)
COTS	Commercial Off-The-Shelf
DDS	Data Distribution Service
DSTO	Defence Science and Technology Organisation
ENGAGE	Extensible Natural Gesture Animation Generation Engine
FOCAL	Future Operations & Analysis Laboratory (DSTO Edinburgh)
HiOP	Higher-level Operating Picture
HTML	Hyper-Text Markup Language
IMMP	Intelligent Multi-Media Presentation
JOAD	Joint & Operations Analysis Division, DSTO.
JOGL	Java OpenGL bindings
MHEG	Multimedia and Hypermedia Experts Group
RST	Rhetorical Structure Theory
SMIL	Synchronised Multimedia Integration Language
SMPL	Simple Multimedia Presentation Language
TCP	Transmission Control Protocol
THML	Talking Head Markup Language
TTCP	The Technical Cooperation Panel (AS/CA/NZ/UK/US)
TTS	Text-To-Speech (system)
URI	Uniform Resource Identifier
VA	Virtual Adviser
W3C	World Wide Web Consortium
XSLT	eXtensible Stylesheet Language Transformation
XML	eXtensible Markup Language

1. Introduction

Situational awareness (Endsley, 1995) is a key requirement for military command and control, allowing military staff and commanders to make the right decisions at the right time. To achieve this they need to discover, understand, and reason with an exponentially increasing volume, velocity, and variety (Laney, 2001) of highly dynamic, complex, and time critical information from sensor, defence, and national information networks. Meeting this 'Big Data' challenge will require increasing reliance on automation to assist with the discovery, retrieval, and fusion of contextually relevant information. However, achieving situational awareness also requires automated systems to present the information in a contextually appropriate manner. The Defence Science and Technology Organisation (DSTO) has an ongoing research program to achieve this using a multimedia narrative paradigm modelled on an interactive news service capability (Wark and Lambert, 2007).

Figure 1: An example of DSTO's intelligent, interactive new service capability with a fictitious scenario.

This presentation paradigm is based around animated characters, dubbed 'Virtual Advisers' (Lambert, 1999, Nowina-Krowicki et al., 2011, Taplin et al., 2001) that provide narrative, and if appropriate, 'expert' commentary on multimedia content. Multimedia content can include text, images, videos, graphs, diagrams, 2D/3D animations, or geospatial scenes. The Virtual Advisers in this case help change, establish, and maintain context so as to facilitate comprehension and projection – they establish 'the story behind the data'. While this can well be achieved using human narration in a briefing role, Virtual Advisers potentially provide an automated capability that can be accessed on demand using the most up-to-date information available. When coupled with a geospatial display,

this can provide a Higher Level Operating Picture (HiOP) that supports perception, comprehension, and projection across the strategic, operational, and tactical levels of command.

There is significant development effort still required to fully realise this capability:

a) Natural language processing and 'understanding' of documents and information sources normally intended for human consumption.

b) Contextually appropriate machine interpretation and representation of multimedia content.

c) Machine reasoning systems to automatically fuse all-source data, identify a developing situation, and extrapolate to the consequences on command intent.

d) Intelligent Multi-Media Presentation systems (IMMP) to automatically assemble appropriate multimedia content, generate and present a narrative from this machine representation that is both informative and engaging, taking into account any limitations or constrains on the rendering system.

e) Automated generation of non-verbal behaviours for animated characters so that generated text is conveyed in an appropriate and engaging manner. Unlike the entertainment industries, in this context manual generation and tuning of behaviours tailored to suit a particular narrative sequence is not appropriate. The current capabilities and limitations of the Virtual Adviser technology strongly influence the nature of the multimedia content that needs to be generated. In particular, limitations of commercial text-to-speech (TTS) systems constrain the vocal inflection and prosody that can be applied to a narrative, requiring other strategies to maintain user engagement.

Multimedia narrative can be considered to be an extension of narrative. Bal's model for narrative structure (Bal, 2009), as adapted by Bui et al (Bui et al., 2010), has three layers of abstraction:

1. *The Fabula*, which represents the narrative environment, events, actors, and beliefs, desires, and intentions of these actors.
2. *The Plot* (or *Story*), which represents a subset of the *Fabula* presented from the point of view of one or more *Focalizers* (or agents) within the narrative.
3. *The Presentation* (or *Text*), which represents how the *Plot* is presented to the audience.

In this model, the generation of the *Fabula* layer is supported by a-c) above, the generation of the *Plot* layer is supported by d) above, and the generation of the *Presentation* layer is supported by d) and e). The work described in this report is focussed on the *Presentation* layer, and on the feasibility of generating different presentations from existing multimedia content (i.e. *Plot*) to suit different audience requirements, using Rhetorical Structure Theory (Mann and Thompson, 1988, Taboada and Mann, 2006) to annotate the multimedia content. This use case could have immediate applicability to support adaptable re-use of any manually generated presentations, such as a 'Road to War' briefing. For this work, an example is chosen based on the Military Strikes in Atlantis Scenario (Blanchette, 2005) as presented during a series of demonstrations developed by the Intelligence Processing and

Analysis Branch of DSTO's Command, Control, Communication and Intelligence Division (now part of the National Security, Intelligence, Surveillance, and Reconnaissance Division) in 2012-13. In this case, the *Plot* layer has been manually generated, but the techniques discussed could also apply to automatically generated narrative *Plots.* Work on automatic generation of a narrative *Plot* from an existing *Fabula* is discussed elsewhere (Dall and Donnelly, 2014).

2. The Virtual Adviser

Virtual Advisers (VA) were developed by DSTO to enable the presentation of multimedia narrative by providing a story telling interface. They are computer generated characters using photo realistic textures with real-time animation and commercial-off-the-shelf TTS generation. Virtual Advisers can include rolling text captions and multimedia monitors à la television news services. Virtual Advisers have been designed for modularity and can be delivered in a number of ways to users.

Virtual Advisers have been used to present situation briefs and other prepared presentations incorporating other media such as tables and diagrams, images, video, web pages and so on. They have also been used to present dynamically generated content incorporating a dialog management system with a conversational interface (Estival et al., 2003). When connected to a decision support system they can also alert people to new or changing situations (Lambert, 1999, Wark et al., 2003).

Virtual Advisers can augment human support staff by providing a capability that can be deployed and accessed simultaneously from multiple geographically distributed locations. They can present the same information repeatedly, on demand, and without imposing an additional manning burden.

Virtual Advisers can be delivered in several ways: as a Desktop Application that can be controlled via an integrated Desktop Service; embedded as an overlay in commercial off-the-shelf (COTS) applications (Kohler et al., 2013), or as a 3D model inside DSTO's Virtual Battlespace (Wark et al., 2009); and as an Applet displayed on web pages and integrated into popular wiki systems, such as Atlassian's Confluence[1] and the ubiquitous, open source MediaWiki[2].

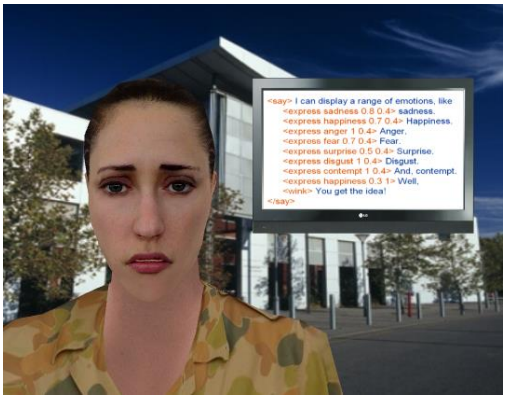
In the following sections we will describe how the Virtual Adviser system is implemented and how it can be integrated with other systems, such as an IMMP system.

2.1 Talking Head Markup Language (THML)

Content is provided to Virtual Advisers in the form of Talking Head Markup Language (THML). THML is tagged text that describes what the VA is to say and do. It includes commands to direct the VA: to say text; to adopt degrees of fundamental facial expressions (happy, sad, angry, afraid, surprised, contempt, disgust) (Ekman and Friesen, 1977) to make eyebrow and head movements; and to direct gaze. It also includes commands to control the underlying TTS system, the appearance of the VA and its environment, and

[1] See https://www.atlassian.com/software/confluence for information on Confluence, or http://logwiki.dsto.defence.gov.au/display/va/Adding+a+Virtual+Adviser+to+a+Wiki+Page for an example.
[2] See http://www.mediawiki.org/wiki/MediaWiki

synchronise with other applications. THML is not a strict XML format, as it was designed to be simple for humans to read and write and to support on-the-fly authorship.



```
<va viewer monitor.texture=http://logwiki.dsto.defe
nce.gov.au/download/attachments/37389119/thml.png>

<say> Virtual advisers respond to marked up text
which identifies what content is to be said in
addition to how that content is to be emotionally
expressed. </say>

<say> Using real-time animation and text-to-speech
generation, I can display a range of emotions,
like: </say>
<say><express sadness 0.8 0.4> Sadness. </say>
<say><express happiness 0.7 0.4> Happiness. </say>
<say><express anger 1 0.4> Anger. </say>
<say><express fear 0.7 0.4> Fear. </say>
<say><express surprise 0.7 0.4> Surprise. </say>
<say><express disgust 1 0.4> Disgust. </say>
<say><express contempt 1 0.4> And, contempt. </say>
<say><express happiness 0.3 1> Well, <wink> You get
the idea! </say>
```

Figure 2: Virtual Adviser screenshot of example THML text

Error! Reference source not found. shows an example of a simple THML script. In this example the following THML commands are seen:

- `<va viewer monitor.texture=…>` is a command that specifies the image to use for the embedded multimedia monitor.
- `<say>…<say>` instructs the Virtual Adviser to say the enclosed text.
- `<express sadness 0.8 0.4>` instructs the Virtual Adviser to display the given fundamental facial expression (sadness). The parameters determine the amplitude of the expression (0.8) and the onset time (0.4 seconds).
- `<wink>` instructs the Virtual Adviser to perform a non-verbal action (wink) after the utterance "`Well,`". As in this example, markup embedded within a `<say>` statement will be performed at that point within the utterance.

A more detailed description of THML can be found in Appendix A.

2.2 VA Architecture

Virtual Advisers are implemented using a modular, distributed architecture. The system consists of three core components; a rendering engine, system controller (THConsole), and Text-to-Speech service. Automated behaviour generation can optionally be provided by ENGAGE (see §2.3). The content to be delivered by the VA can either be provided interactively by the user or, more typically, by a content generation system that feeds the THConsole the THML to be presented on demand, such as the IMMP system discussed in this report.

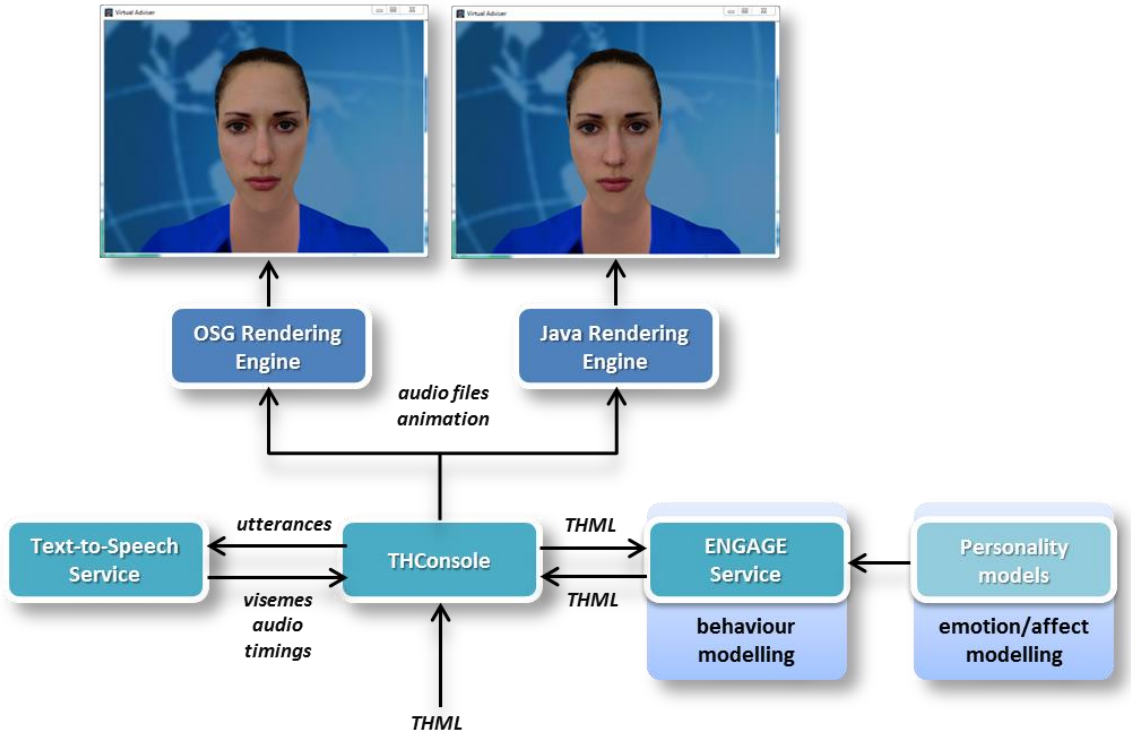


Figure 3: The Virtual Adviser Core System Architecture

In this architecture the rendering engines are used to display the VA to the user. They receive low-bandwidth rendering and timing instructions from the THConsole and output correctly synchronised 3D graphics, audio and external application control. The Virtual Adviser is rendered in a 3D scene, and it uses the CAL3D[3] library to provide character animation. C++ and Java toolkits have been developed to provide reusable, cross platform, core features to help facilitate the rapid development of new rendering engines. These toolkits provide additional common underlying functionality such as: pluggable audio (via OpenAL[4]/JOAL[5]/LWJGL[6] and SDL[7]), instruction parsing, an event based timeline system, and networking support.

The core VA architecture has been implemented in a number of different ways. Two of these implementations are discussed next.

[3] See http://home.gna.org/cal3d/
[4] See http://www.openal.org/
[5] See http://jogamp.org/joal/www/
[6] See http://lwjgl.org/
[7] See https://www.libsdl.org/

2.2.1 Desktop Service Architecture

The Virtual Adviser can be deployed on a Windows or Linux desktop as a service that renders content to the user on demand as show in Figure 4. This implementation uses a rendering engine built with the high performance OpenSceneGraph[8] 3D library, which provides features such as: stereoscopic viewing; integrated video and multimedia; tickertape captioning; an embedded web browser; and Render-to-Texture support.

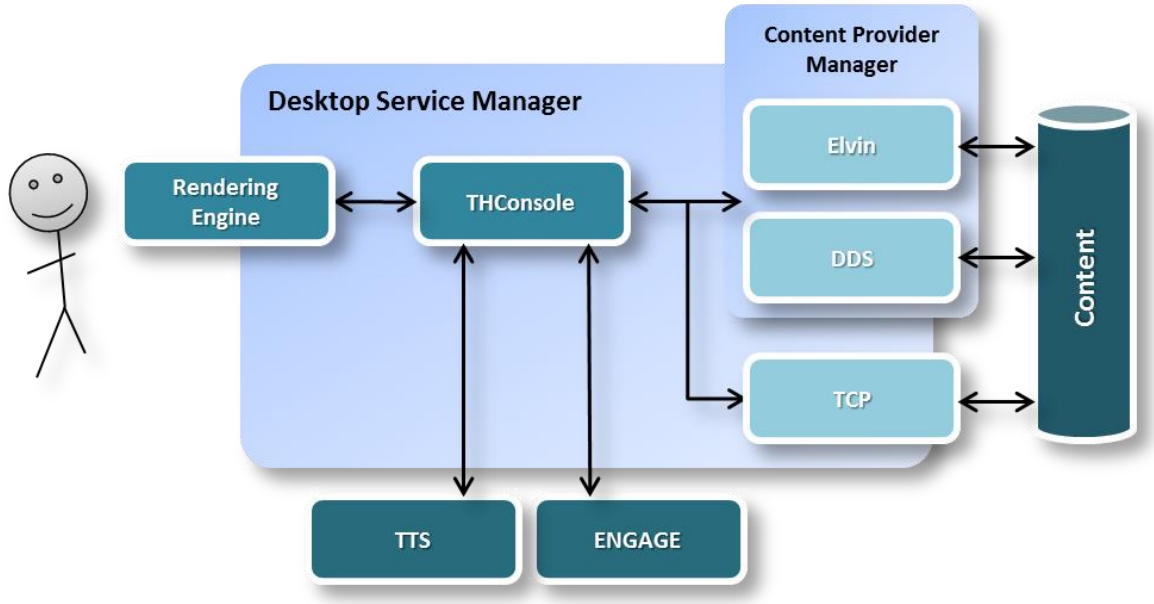


Figure 4: Virtual Adviser Desktop Service Architecture

Content can be provided to the desktop system in one of four ways:

- Interactively by typing THML commands into the control panel.
- A TCP socket connection that accepts THML content.
- As a formatted message using the Elvin Enterprise Bus (or Avis[9]), which has fields for:
    - Priority – priorities between 0 and 1 determine the precedence by which messages are handled; priorities greater than 1 invoke an interrupt.
    - Assigned attribute-value pairs – subscription conditions can be used to determine which messages are handled.
    - Text – provides the utterance for the Virtual Adviser. This can also include embedded THML commands.

[8] http://www.openscenegraph.org/

[9] Avis is an open-source implementation compatible, see http://avis.sourceforge.net/index.html

- A Data Distribution Service (DDS)[10] subscriber that can listen on a nominated topic.

The last three of these mechanisms could be used by an IMMP system.

2.2.2 Web Service Architecture

The Virtual Adviser can also be deployed as a Web Service that is displayed via a Java based Applet using the Java OpenGL (JOGL[11]) bindings. The Applet works on all major platforms and across all web browsers that support the Java plug-in. A JavaScript API allows the Virtual Adviser to interact with the page content and be dynamically controlled using AJAX. Wiki integration for Confluence and MediaWiki allow users to easily embed and control a Virtual Adviser on a wiki page.

This can also be integrated into an IMMP system. For example, a PowerPoint presentation capability has been implemented in Confluence that exploits this, as shown in Figure 5.

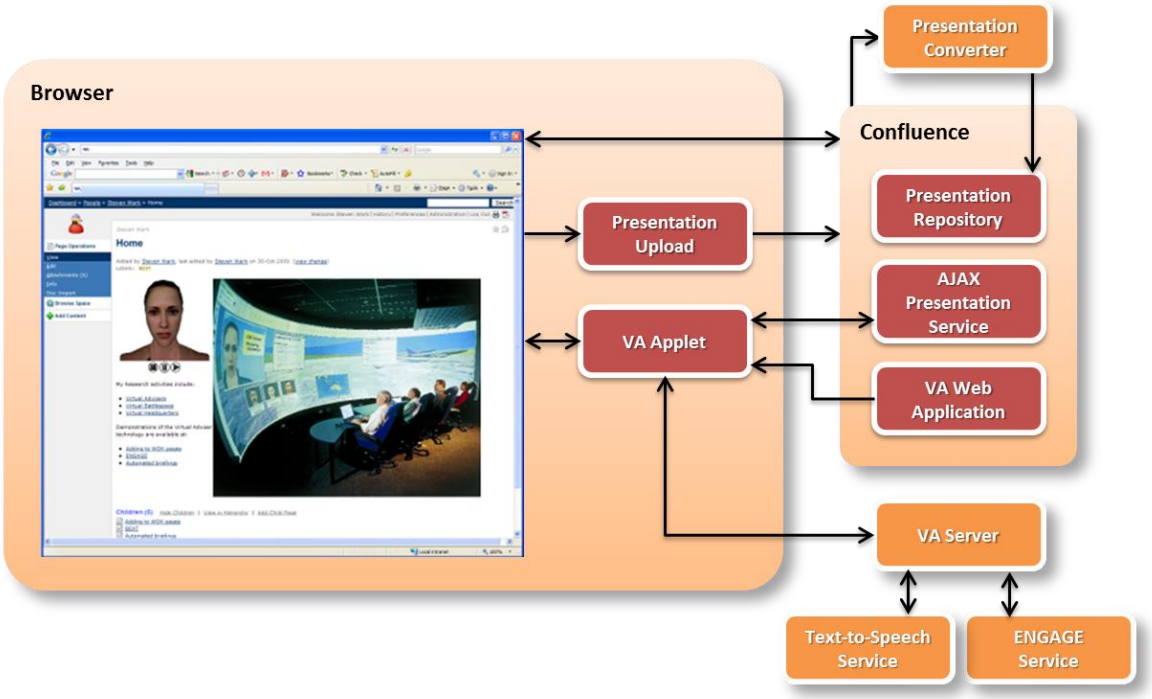


Figure 5: Virtual Adviser Web Service Architecture as implemented in Confluence to provide a presentation capability.

[10] Using the RTI DDS implementation, see http://www.rti.com/products/dds/
[11] See http://jogamp.org/jogl/www/

2.3 ENGAGE

The Extensible Natural Gesture Animation Generation Engine (ENGAGE) provides the Virtual Adviser system with automated character behaviour generation based on sentence syntax (Nowina-Krowicki et al. 2011). ENGAGE takes care of the intricacies of character animation, leaving the content author to focus on the subject matter at hand.

ENGAGE has been designed with the following objectives in mind:

a) Enhance user engagement with Virtual Advisers

b) Let content authors focus on content, not animation

c) Support the comprehension of information through the display of:

- Confidence
- Urgency
- Importance

d) Manage content in a context sensitive manner

- e.g. Abbreviations

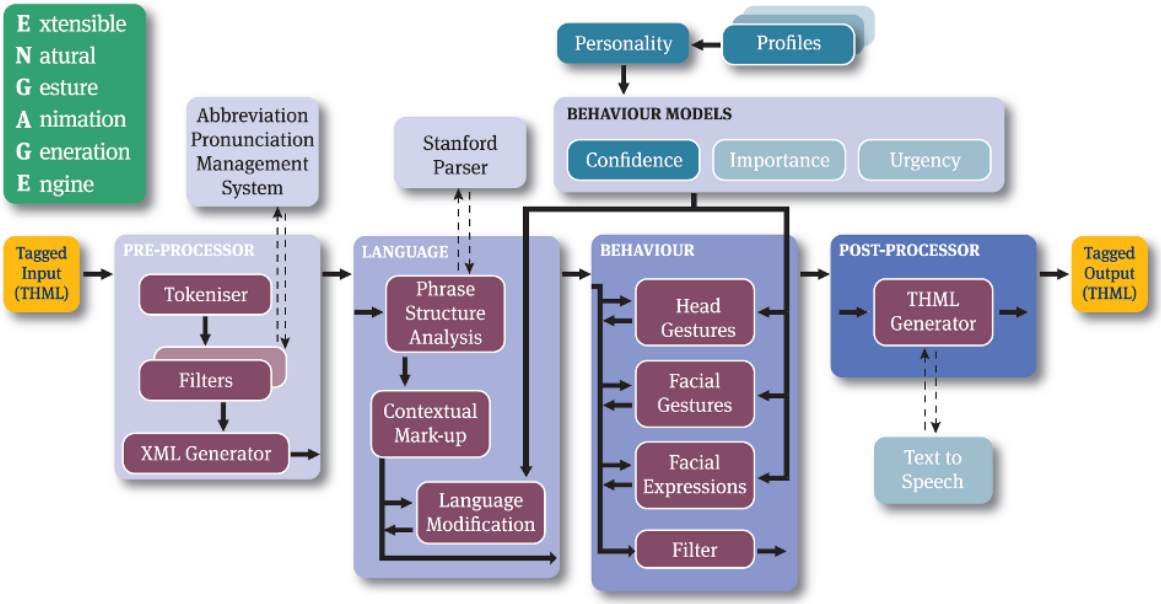


Figure 6: The ENGAGE System Architecture

ENGAGE uses a pipeline to augment THML with appropriate behaviour and language modification as shown in Figure 6. The pipeline consists of the following stages:

1. Input to the ENGAGE system is THML that optionally includes tags to specify the confidence, importance and urgency associated with the content. In addition optional context tags can be included that describe the application domain for the content.

2. The pre-processor prepares the content for language markup by converting the input into XML for further processing. This includes appropriately handling the content based on the context tags, for example how to expand an abbreviation or pronounce a domain specific term.
3. The language module takes the surface text, i.e. the text that you want the Virtual Adviser to say, and automatically marks it up with syntactic information using the Stanford Parser[12]. The behaviour model can modify the language at this point, for example low confidence may result in slower speech and the insertion of speech disfluencies such as hesitations, "umms" and "errs".
4. The behaviour module inserts appropriate non-verbal behaviours, such as head nods, eyebrow raises, eye movements and facial expressions. The behaviour model can modify the animation at this point, for example low confidence may result in a change of facial expression, a head slump and gaze drop.
5. The post-processor takes the generated XML and converts it back to THML for rendering by the Virtual Adviser.

The context, confidence, importance and urgency tags influence how the Virtual Adviser presents the content provided to it. These tags could also be used in a similar way in a more general automated IMMP system. The ontologies developed for these should also have use in an IMMP system.

2.4 Multimedia Narrative using Virtual Advisers

2.4.1 Multimedia Capabilities

Virtual Advisers provide several presentation modes, in additional to the character animation and vocalisations, that support multimedia narrative. These include:

- A choice of character and clothing – usually used to set the context of the presentation and associate a particular character with a particular domain of expertise.
- A background scene that can be a graphic, image, video, or animation – usually used to set the context of the presentation.
- One of more multimedia monitors (and/or overlays), à la television news services, that supports the display of:
    - 2D graphics and images – which can include, for example, content from a PowerPoint presentation.

[12] See http://nlp.stanford.edu/software/lex-parser.shtml

  - Video and animation – including the associated audio, using all of the common formats supported by FFmpeg[13].
  - An embedded web browser – supporting, for example, dynamically generated content provided by web services.

- One or more text captions – supporting different fonts, text colours, text sizes, and an option for rolling 'ticker-tape' captions.
- One or more caption icons – images and/or graphics that can be associated with a block of text in a caption. For example, this might be used to indicate that a particular message is important, or is a routine status update.
- Playback of audio files - as part of the 'scene' or as part of the presentation.
- One or more 3D models – can be inserted in the Virtual Adviser's 3D scene to provide, for example, set props (e.g. a desk or podium) to help establish the context, or media as part of the presentation.

These features allow the Virtual Adviser to provide a multimedia narrative in a similar way to a human presenter.

2.4.2 Content Capabilities

THML includes a number of features to allow the content of a presentation to be separated from the layout of the presentation material in the 3D scene. This allows a modular and flexible approach to content creation and rendering. These features include:

- **Macros** – macros allow variable substitution in content provided to the Virtual Adviser, thus providing dynamic evaluation at run-time. In particular, this allows abstract references to features of the Virtual Advisers 3D scene to be used in a presentation, with realisation at run-time.
- **load** – allows THML content to be included at run-time from a file or web resource. This allows a modular approach to content generation and storage.
- **wait** – provides a mechanism to synchronise multiple concurrent content. For example this would allow the Virtual Adviser to wait for a video to complete before continuing.
- **groovy** – allows a Groovy script to be executed to dynamically generate content based on the values of macros and the parameters passed to the script. This has clear benefits for an IMMP system.
- **loadxml** – allows an XML document to be loaded and transformed, using one or more specified XSLT transforms to generate THML. This provides interoperability with the W3C and other standards.

[13] See http://en.wikipedia.org/wiki/FFmpeg

- **xml** – this is similar to the loadxml feature, but allows for the XML to be embedded within the THML prior to transformation. This provides additional processing options.
- **script** – this provides a generic scripting interface that allows any embedded text to be transformed by a scripting language into THML. This extends the capabilities of the xml feature.

An automated IMMP system could dynamically generate THML content on the fly to suit the presentation requirements. The features discussed here facilitate human authoring and reuse of content, and have been demonstrated, through our previous work, to provide a useful capability for this. This has motivated the current work into a more flexible multimedia presentation capability exploiting these features.

2.4.3 Limitations

While Virtual Advisers have been developed to provide a 'natural' story-telling interface for multimedia content, they are still a long way from being able to replicate the capabilities of a (good) human story-teller. Key to this is the ability to engage and focus the attention of the audience. The technologies currently used with the Virtual Adviser system have limitations that strongly influence what it is capable of with regards to:

- **User Interactivity:** with appropriate speech recognition and dialog management systems, a Virtual Adviser can respond to simple requests or commands from a single user, but it cannot, for example, interact extemporaneously with an audience. Thus, it cannot deal well with interruptions from an audience, except via a mediator, nor can it recognise when it is interrupting a human user.
- **Spatial Awareness:** in most use cases considered, the audience does not participate in the 3D virtual world which the Virtual Adviser inhabits. This has the consequence that the Virtual Adviser cannot direct its attention, or its presentation, to individuals in the audience except via verbal cues. All audience members see the same thing, and so the Virtual Adviser cannot engage a particular individual by looking at them. Coupled with the limitations of verbal interactivity discussed above, this places severe limits on how the Virtual Adviser can interact with individuals in the audience – except via a mediator. Interacting with the audience is a powerful means of user engagement available to a human presenter that is not easily available to a Virtual Adviser. However, Virtual Adviser can, for example, look everyone in the eye at the same time, so other approaches to user engagement may be available.
- **Pronunciation:** most commercial-off-the-shelf TTS systems now available can do a reasonable job of annunciation of the text provided to them, with few pronunciation errors. However, those pronunciation errors that do occur are often jarring and tend to disrupt user engagement. It is for this reason the Abbreviation and Pronunciation Management System (APMS) has been included in ENGAGE – but to be effective, the author (or automated generator) of any content for presentation needs to ensure that the correct pronunciation is used for the context of the presentation.

- **Verbal Expressivity:** most commercial-off-the-shelf TTS systems now available provide a simple level of vocal inflection in the generated speech that is appropriate for most cases. However, there is limited fine-grained control over the speech generation so that, for example, the Virtual Adviser is not able to sound sad if it is looking sad, and it is not able to emphasise particular parts of its utterance. This tends to make the Virtual Adviser appear to 'drone' on for long utterances. Unlike a good human story-teller, it is not able to actively engage the audience through modulation of its voice. For this reason, it is advisable to avoid monologues and keep the Virtual Adviser utterances short (~30-60 seconds), breaking them up with other content or actions. The presentation needs to be cast so that any long extracts of text should be presented in some other way, not as an utterance. Similarly, numbers more than 2 or 3 digits should be displayed rather than spoken. It is also useful to reinforce any longer utterances with a short text caption.
- **Verbal Disfluencies:** 'umms', 'arrhs', coughs, and other sound effects are often used, either consciously or unconsciously, as part of a presentation and convey cues about the content being presented – whether it is the degree of confidence they have in the material, or it is part of the material itself. There is limited (and often none) support for these sorts of disfluencies in almost all commercial-off-the-shelf TTS systems.
- **Behaviour Models:** most of us spend a lifetime learning the appropriate social behaviours for the wide range of circumstances we find ourselves in, and are able to rapidly adapt these as needed. While the ENGAGE system we have developed will allow us to prescribe certain behaviours for the Virtual Adviser *with appropriately marked up content*, only a very limited set of behaviour models are currently implemented. This means that the non-verbal expressivity of a human presenter is not easily achieved – it would require fine-grained mark-up of the content, which is an unreasonable expectation for most content authors. This needs to be considered when creating content for a Virtual Adviser presentation.

Arising from these limitations, some simple guidelines for multimedia presentations with Virtual Advisers are:

- **Keep It Simple:** this is particularly true for content intended as utterances, in order to avoid any situations where the TTS may have trouble generating the speech. Any complex or contextually dependent terms should be pre-defined in ENGAGE's Abbreviation and Pronunciation Management System.
- **Keep It Short:** try to keep utterances to less than ~30-60 seconds. This means that other media or modes should be used to break up the utterances in a presentation.
- **Keep It Sweet:** the Virtual Adviser will continue on with its intended presentation with single-minded determination – make sure that the content is at the appropriate level for your audience, as the Virtual Adviser is very bad at ad-libbing.

- **Say It Three Times:** it's important to try to reinforce (not repeat) content presented, usually through summarisation of some form.

Not surprisingly, these guidelines are not dissimilar to those given to a human presenter, particularly an inexperienced human presenter, for similar reasons.

3. Intelligent Multi-Media Presentation (IMMP)

Intelligent Multi-Media Presentation systems automate the selection, design and presentation of multimedia content. The advantages of IMMP systems are recognised to be (Andre, 2000, Paris et al., 2004):

- Adaptability and flexibility provided by generating multimedia presentations on-the-fly from available information and content to suit a particular user or audience in a particular situation.
- Coherence within a presentation by maintaining consistency across the information and content used within the presentation.
- Effectiveness by designing presentations that take into account the characteristics of the information sources, the task that the users need to perform, and the communicative goals to be achieved.

Previous work on IMMP systems in DSTO has looked at, in collaboration with CSIRO, a framework for an IMMP system within the Future Operations Command Analysis Laboratory (FOCAL) (Colineau and Paris, 2003, Paris et al., 2004), and implementation of a prototype system (Andrews et al., 2006) using the ATTITUDE multi-agent system (Lambert, 1999). More recently, work has been done on a conceptual design for IMMP using model-based systems engineering techniques (Nugent, 2012), which includes a review of IMMP systems that have been implemented.

In this section we will review some of the considerations for an IMMP system, and describe a proposed system for authoring of 'intelligent' multimedia presentations for Virtual Advisers, as a step towards an automated news service for military situational awareness. In this case, the aim is to provide a system that provides adaptable and flexible multimedia presentations that suit a particular audience and situation.

3.1 Background

Multimedia generation poses some unique challenges, such as how to tailor and coordinate text and graphics to complement each other, but there is also much similarity to the challenges posed by natural language generation. Consequently, the techniques used for automated multimedia presentation have drawn heavily on the lessons learned during the development of systems for the automated natural language generation for textual discourse (Andre, 2000).

3.1.1 Communicative Goals

Generalising the approaches taken with natural language generation, the generation of multimedia presentations has been treated as a goal-directed activity by many researchers (Paris et al., 2004). A communicative goal is used to build a multimedia presentation, structured as a hierarchy of communicative acts, each supporting a specific sub-goal. For example, a presenter may point to an illustration or animation while providing a

commentary, to achieve a specific sub-goal within the presentation that contributes to the intent (goal) of the whole presentation.

Creation and presentation of multimedia material, or re-using content in another context, can both be considered as communicative acts within different types of multimedia presentations (Andre, 2000):

1. Multimedia content is generated and used at the same time – for example, when a presenter draws on a blackboard and provides commentary.
2. Multimedia content is created and re-used at a later time by the same person – for example, when someone prepares in advance the content for a presentation.
3. Multimedia content is created and re-used by a different person – for example, when someone reuses material from another source.

In the last two, the goals underlying the production of the material may be quite different from the goals to be achieved by presenting it. The IMMP system needs to support these different interpretations, and this provides part of the motivation behind the structure of the multimedia presentations discussed later in this report.

3.1.2 Coherence

The coherence of a text discourse, or, in general, a multimedia presentation, describes how well the individual communicative acts contribute to the communicative goal. Coherence requires understanding of the relationships between elements of a discourse or multimedia presentation, of how these are aggregated to form larger discourse or presentation elements, and finally how these discourse or presentation elements are organised. Coherence often depends on the structure of a discourse or presentation and how well it adheres to an expected schema, which depends on the topic and context of the presentation. For example (Colineau and Paris, 2003), you may expect to see a restaurant menu structured into courses: Entrée; Main; Dessert; and Beverages. The placement of each dish within the menu establishes its role within a meal, and helps maintain the coherence of the whole menu. Dishes appearing in the wrong section would break the menu's coherence, and so would dishes grouped into unfamiliar categories.

The use of a schema to determine the selection and organisation of text for text-based systems has proven to be an effective way of selecting and organising content that maintains coherence (McKeown, 1985). This approach has been generalised and applied to the organisation of multimedia presentations, and the structure of the textual components within them.

The structure of a discourse or presentation can be characterised by the hierarchy of so-called rhetorical relations between its elements. Coherence can generally be achieved when an appropriate hierarchy of rhetorical relations applies to a discourse or presentation. One of the most elaborate and commonly used sets of rhetorical relations for discourse and presentation generation is Rhetorical Structure Theory (RST) (Mann and Thompson, 1988), which will be described in detail in a later section. The structural schema appropriate for a particular application domain can usually be formed primarily

from the RST relations, with a few specialised relations added to support the particular domain.

3.2 Multimedia Design

Multimedia presentation systems face additional challenges to text-based systems, including:

- How to find a media combination that conveys the communicative goal effectively in a given situation.
- How to distribute and coordinate different media onto different renderers.
- How to tailor the different media so that they can be presented together without distracting the user/audience.
- How to integrate the different media so that they convey the communicative goal.

3.2.1 Terminology

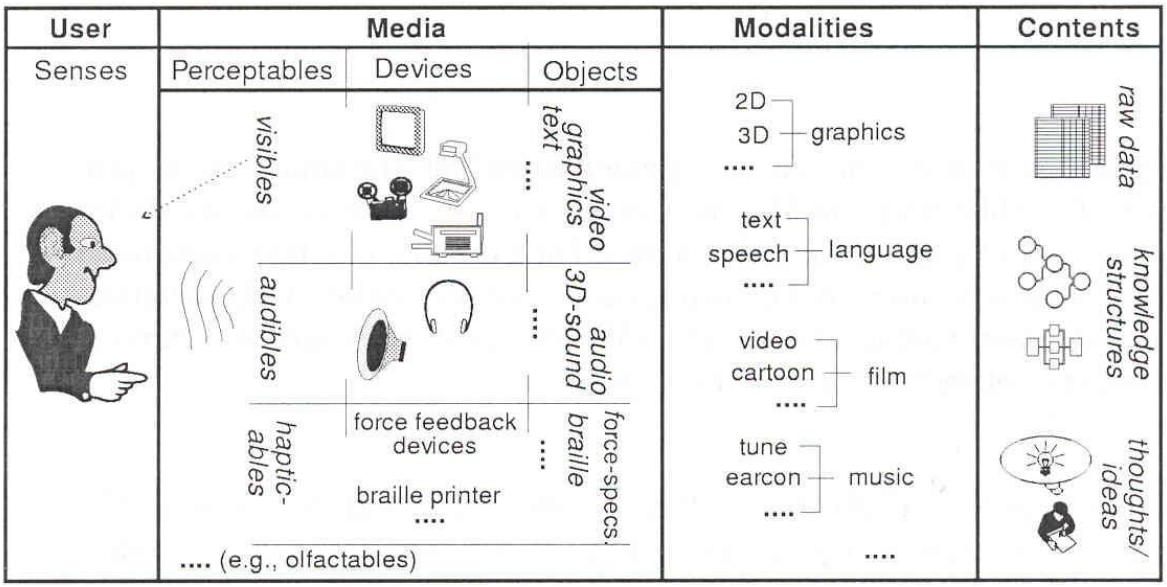


Figure 7: The distinction between the medium and modality (Bordegoni et al., 1997).

Applying the terminology of Bordegoni et al. (Bordegoni et al., 1997):

- *Medium:* refers to the perceptual channel, the physical devices that provide information in this perceptual channel, and the 'type' of information presented – it is closely tied to the sensory and cognitive processing capabilities of the user/audience;
- *Modality:* refers to the way the information is encoded within a particular format. Note that modality is not orthogonal to medium – for example, language can be

presented as written text in one medium, or in a different modality (and medium) as speech.

With this terminology, a multimedia presentation could also include information in multiple modalities. For example, a television news report including commentary from a reporter, a photograph, and a caption contains multiple media (graphics, text, audio) and multiple modalities (text and speech). In contrast, a multimodal presentation could include only a single type of information, but encoded in different ways – for example, a text document including bullet points and tables could be considered as a multimodal document, but not a multimedia document. With our focus on multimedia narrative, we expect that multimedia presentations of interest to our discussion will also include multiple modalities.

3.2.2 Media Allocation

The selection of the media, and the modality used, in a multimedia presentation is influenced by several factors (Andre, 2000, Colineau and Paris, 2003):

- **The characteristics of the information to be conveyed:** different types of information have been found to be conveyed more effectively by different media. For example:
    - Graphics is preferable for conveying visual information such as relative size, shape, colour, texture.
    - Graphics is preferable for spatial or temporal relationships such as relative location or orientation.
    - Text is preferable where accuracy of spatial or temporal relationships is important, such as spatial dimensions or exact coordinates are required.
    - Text is preferable for conveying linear or causal sequences
    - Text is preferable for qualitative information such as: most, some, any, exactly.
    - Items that are contrasted with each other should be presented in the same medium.
- **The communicative goal:** the selection of media and modality clearly depends on the context in which the presentation is given, and the communicative goal to be achieved. Different media and modalities may be more effective in different situations.
- **The user's characteristics:** Different users may have different information processing styles, and be better able to comprehend information presented in different media and modalities. Furthermore, different audiences may have different expectations for the schema of a multimedia presentation, which may well influence the media and modalities chosen.
- **The combination of modalities:** The combination of several media or modalities is most effective when these media/modalities are integrated so that each medium/modality contributes to the understanding of the whole presentation.

This can be achieved, for example, by using co-references between media elements, and spatial or temporal contiguity of related information.

- **The resources and available media:** Resources may impose constraints on the way that information can be presented, and hence on the selection of media for the presentation. For example, if an audio capability does not exist, then aural media is inappropriate.

3.2.3 Cohesion

As discussed above, the combination of multiple media/modalities into an effective multimedia presentation requires more than just the simple juxtaposition of multimedia content, but the integration of each element into the presentation to reinforce the cohesion between these elements. Research into how to achieve this has shown that in a multimedia presentation the following types of referring expressions can be applied to maintain cohesion (Andre, 2000):

- **Multimedia referring expressions:** refer to objects using a combination of two or more media, each of which conveys some discriminating attributes that need to be taken together to provide a complete reference. For example, the utterance "located here on the map" while pointing at an object on a map.
- **Cross-media referring expressions:** refer to other elements in the multimedia presentation. For example, the text "as shown in Figure 1". In most cases, cross-media referring expressions serve to direct the audience's attention to a particular element in the multimedia presentation that needs to be interpreted to convey the communicative goal.
- **Anaphoric referring expressions:** refer to objects in an abbreviated form, assuming they have already been introduced, either explicitly or implicitly. For example, in the text "Jane is a Virtual Adviser. Her head is pointy", "Her" is an anaphoric reference to the antecedent "Jane".

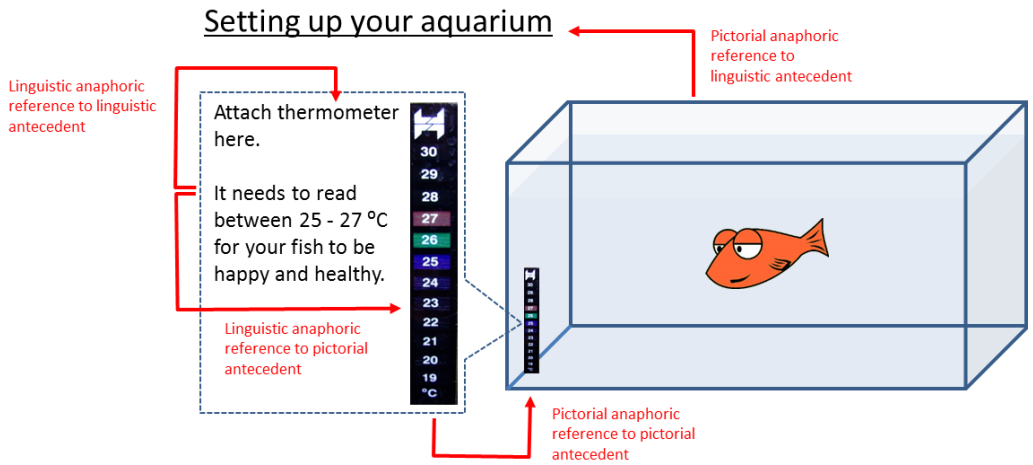


Figure 8: Examples of multimedia anaphoric references.

In multimedia presentations there are several forms of anaphoric references possible:

a. Linguistic anaphora with pictorial antecedents – for example, in Figure 8, "It" provides a linguistic anaphoric reference to the graphic of the thermometer.

b. Pictorial anaphora with linguistic antecedents – for example, in Figure 8, the diagram of the aquarium provides a pictorial anaphoric reference to the linguistic description of the diagram.

c. Pictorial anaphora with pictorial antecedents – for example, in Figure 8, the enlarged view of the thermometer provides a pictorial anaphoric reference to the diagram of the aquarium.

In some sense multimedia anaphoric referring expressions can be considered to be special cases of multimedia referring expressions, where the different media elements provide shorthand representations for the others. Also, while an anaphoric reference can be considered as a binary expression, a multimedia referring expression can have larger cardinality.

Illustrations are often incorporated into referring expressions, as they provide a focus for the content of the presentation and provide a ready means of discriminating between alternatives. Within a multimedia presentation, the features of the illustration may be referred to, as well as the features of the object depicted. Thus, it must be clear whether the presentation is referring to the features of the illustration, or of the object depicted.

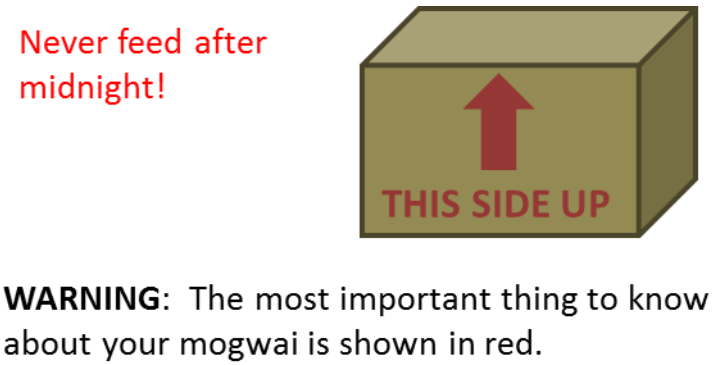


Figure 9: An example where it is not clear whether the text is referring to the features of the object depicted in the illustration, or the features of the illustration.

Spatial relationships are often used to discriminate between different multimedia elements – for example, "the diagram above shows how referring expressions can be misinterpreted". In general, it is not possible to know beforehand the layout of a multimedia presentation, or how it will be rendered, so that these referring expressions will either need to be generated dynamically at run-time, or be provided by an indirect reference to the content – for example, "as shown in the monitor", where the location of the monitor is established beforehand either explicitly or implicitly. This is also the case when 3D content is being presented – the viewing angle and position may change (for

example, if it is controlled by the audience), and so even references to components a single multimedia element may need to be handled with care.

3.2.4 Multimedia guidelines

Studies in multimedia learning have established some guidelines that could be applied to multimedia presentations (Mayer and Moreno, 2003). This work was focussed on how well knowledge obtained from multimedia explanation of causal systems (using animation, on-screen text, and narration) was transferred to problem solving, but it may be useful to consider these guidelines in a broader context. This work supported a model for multimedia learning based on three assumptions:

1. **Dual Channels:** humans possess separate information channels for verbal and visual material
2. **Limited Capacity:** There is only a limited amount of processing capacity available in the verbal and visual channels.
3. **Active Processing:** Learning requires substantial cognitive processing in the verbal and visual channels.

Based on this model, and studies done under situations of various types of cognitive overload, nine strategies for improving knowledge transfer in multimedia learning were hypothesised and validated:

1. **Modality[14]:** Better knowledge transfer occurs when words are presented as narrative than as on-screen text, as this engages both the verbal and visual channels.
2. **Coherence:** Better knowledge transfer occurs when extraneous material is excluded, as processing of extraneous materials uses cognitive resources.
3. **Signalling:** Better knowledge transfer occurs when signals are included in presented material to highlight key content, allowing processing resources to be targeted at this content.
4. **Spatial Contiguity:** Better knowledge transfer occurs when text is placed near corresponding parts of graphics, to reduce processing required for scan/search of content.
5. **Redundancy:** Better knowledge transfer occurs when words are presented as narration only, rather than as narration and on-screen text. Use of on-screen text when it is not required uses processing capacity for the visual channel unnecessarily.
6. **Segmentation:** Better knowledge transfer occurs when the lesson is presented in user-controlled segments than as a continuous unit, to allow them to match information rate to their processing capacity.

[14] Modality in this case referred to the sensory channel exploited.

7. **Pre-training:** Better knowledge transfer occurs when students already know names and behaviours of system components, so that they spend more of their processing capacity understanding the causal relationships of the content.
8. **Temporal Contiguity:** Better knowledge transfer occurs when corresponding animation and narration are presented simultaneously rather than successively.
9. **Spatial Ability:** High spatial learners[15] benefit more from well-designed instruction than do low spatial learners.

In multimedia presentations where more elements are available than the three studied in this work or in situations other than explanation of causal systems, some trade-off amongst these strategies may be required. For example:

- On-screen text, rather than narration, may be preferable for conveying long lists where the temporal contiguity between the first and last elements may be lost using the verbal channel.
- On-screen text, rather than narration, may be preferable for conveying precise numerical information, where temporal contiguity between the first and last digits may be lost using the verbal channel.
- On-screen text may provide a useful way of signalling, as information is presented using narration. Note in this case, it is important to keep the text succinct to avoid overloading the visual channel.
- Presentation of background material may serve a 'pre-training' function so that the core of the multimedia presentation can be better understood.

3.2.5 Layout

As discussed in the previous section, the spatial and temporal layout of the multimedia elements has an impact on the effectiveness of the multimedia presentation at achieving its communicative goal. However, the options available for a multimedia presentation depend on the rendering and display environment. An IMMP system needs to manage and adapt the layout of multimedia elements to different display and rendering environments in order to maximise its effectiveness. This could be as simple as selecting an appropriate presentation template, or as sophisticated as automatically calculating the optimal spatial and temporal layouts at run-time.

3.2.5.1 Presentation Arrangement

The considerations for the arrangement of multimedia elements in a presentation are (Colineau and Paris, 2003):

[15] In this context, Mayer and Moreno defined 'high-spatial' learners as those with the ability to hold and manipulate mental images with a minimum of mental effort.

- **Grouping:** the user's understanding of the presentation is enhanced by grouping closely related material together.
- **Placement**: influences what elements are seen first and last, what the purpose of the content is, and what is of primary or secondary importance.
- **Alignment:** contributes to the legibility and ease of understanding of the whole presentation. This includes attributes such as font and image size, etc.

By adhering to schema appropriate for the context of a multimedia presentation, the grouping, placement, and alignment of content helps establish the structure of the presentation and the rhetorical relationships between the multimedia elements. For example, the arrangement of text and images in a newspaper article establishes which element is the headline, which is the image caption, and which is the body of the article.

3.2.5.2 Presentation Scheduling

An IMMP system also needs to manage the temporal coordination of multimedia elements in a multimedia presentation, again according to a schema appropriate for the context of the presentation. The synchronisation of multimedia elements usually involves the following phases (Andre, 2000):

1. High-level specification of the temporal behaviour of a presentation, usually in terms of qualitative and metric constraints. For example, "show this slide before that one", or "the presentation needs to take no more than ten minutes".
2. Computation of a partial schedule at 'compile-time', satisfying predictable temporal constraints that schedules multimedia elements on a time axis. Since some multimedia elements do not have predictable durations, there is some flexibility in stretching or shrinking the time intervals between multimedia elements.
3. Adaptation of the schedule at run-time as unpredictable multimedia elements are realised.

Multimedia elements can have four primary attributes that determine how they are scheduled (Buchanan and Zellweger, 2005):

1. **Granularity:** refers to the amount of internal structure (events) that is accessible to the system and author.
    a. *Coarse granularity:* only the start and end points of the media are known.
    b. *Fine granularity:* the (relative) time of internal events in media are known – for example, the start of a lion's roar in documentary footage.
2. **Duration:** refers to the length of time required to prepare and present a media component. This can include, for example, the time taken to load a video clip as well as the time taken to play the clip. Duration can be classified in two ways:
    a. *Predictable:* the duration is well known prior to presenting the media – for example, the time to play the video clip above.

   b. *Unpredictable:* the duration cannot be reliably predicted in advance of the presentation of the media – for example, the time to load the video clip above from an internet resource.

3. **Flexibility:** refers to attributes that measures how the media duration can be varied. Flexibility can be classified in two ways:
   a. *Continuously adjustable:* specifies a range over which the duration can be varied. For example, a video clip might have a range of 5 to 10 seconds, with a preferred duration of 8 seconds.
   b. *Discretely adjustable:* specifies discrete values that can be selected – for example, a conference presentation may be available in a 5 minute (poster), 15 minute (paper), and 30 minute (plenary) variations.
4. **Flexibility Metrics:** optionally specify metrics that provide a cost function for a media component as the duration is manipulated, to allow the "best" schedule to be automatically generated.

Analogously, the temporal relationships, that describe how multimedia elements can be combined in a schedule, can also be quantified using four primary attributes:

1. **Granularity:** refers to whether temporal relationships can be placed between points in time, temporal intervals, or both.
   a. *Points:* can be an absolute time, a relative time with respect to the start of the presentation, an event in a presentation (e.g. end of a video), etc.
   b. *Interval:* can be the duration of a media element, a portion of a media element, etc.
2. **Temporal Relation Type:** can be grouped into three main classes:
   a. *Ordering Relations:* binary relations that specify the order of occurrence of points or intervals in the document, based on the 13 Allen temporal relations: *before*; *starts*; *finishes*; *meets*; *overlaps*; *during*; their inverses, and *equals*. Different ordering relations apply depending on whether we are talking about points or intervals.
   b. *Duration Relations:* apply between the durations of different intervals, such as *shorter than* or *longer than*.
   c. *Group Relations:* allow intervals and points (and content) to be grouped together so they can be scheduled as a single entity within a presentation.
3. **Flexibility:** can be specified in two ways:
   a. *Priority:* allows the author to specify that some temporal relationships can be ignored if necessary to meet higher priority constraints.
   b. *Range:* specifies a set of times in which a temporal relationship is deemed to be satisfied. For example, if two video clips finish within a second of each other, they may be deemed to have finished together.

4. **Flexibility Metrics:** as with media components, these optionally specify metrics that provide a cost function for flexibility of temporal relationship, to allow the "best" schedule to be automatically generated.

The combination of these multimedia attributes and desired temporal attributes determine how multimedia elements are scheduled.

Separately from the multimedia elements, the multimedia presentation may also require meta-changes to occur at specified points or intervals. These could be changes in the spatial layout, styles, or transitions from one state to another. For example, a fade-out may be required at the end of a video, or a fade-in to a new scene or slide. These changes are strictly not part of the multimedia content of the presentation, but are in how the presentation is displayed. However, they are still an important aspect of the presentation that needs to be considered.

3.3 A Standard Reference Model

A standard reference model for IMMP systems has been proposed to facilitate collaboration between researchers working in this area, arising from the recognition that many different problems need to be resolved to achieve a comprehensive IMMP capability (Bordegoni et al., 1997). The standard reference model allows:

- a uniform approach to be applied to analysis of IMMP systems
- modular development of IMMP capability
- comparison of different IMMP systems
- a common terminology

However, it should be noted that the standard reference model provides a logical view of an IMMP system, not an implementation blueprint.

The standard reference model conceptualises an IMMP architecture into five layers representing particular subtasks: Control Layer; Content Layer; Design Layer; Realisation Layer; and Presentation Layer. These layers can exploit knowledge resources maintaining 'expertise' about the Application, Context, User, and Design considerations.

Presentation goals and application data form the input to the IMMP system. Goals and application data are processed through the layers, possibly with user input, and are formed into multimedia presentations provided to an end user. An IMMP system may also interact with external systems to obtain information needed to generate a presentation, and may also provide outputs to other systems to allow them to exploit IMMP products.

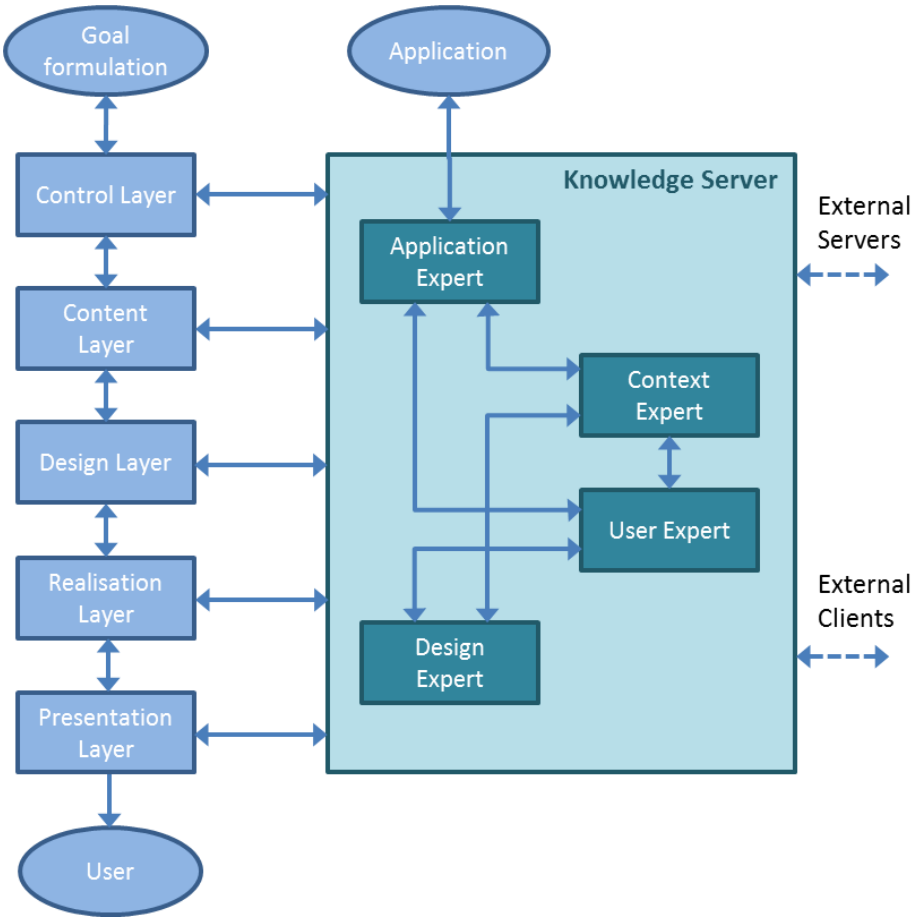


Figure 10: A standard reference model for IMMP [from(Bordegoni et al., 1997)]

3.3.1 Control Layer

The Control Layer fulfils two main interdependent control functions that make use of available knowledge resources:

1. **Goal Formulation Interface:** to allow the user to formulate presentation goals, including selection of available options to refine content generation. This may be as simple as a menu selection, or as complex as a natural language dialogue.
2. **Goal Selection:** to determine, perhaps with user input, what sub-goals to be generated next, and to control execution of the generated presentation (e.g. 'start', 'stop', 'pause', 'back', 'next' commands).

3.3.2 Content Layer

The Content Layer includes four high-level interdependent authoring tasks that utilise available knowledge resources:

1. **Goal Refinement:** this encompasses both the decomposition of a goal into a set of sub-goals, and the specialisation of abstract goals into communicative acts.

2. **Content Selection:** this interacts with the goal refinement process to select the communicative acts, and the relationships between them, that are most appropriate for the application.
3. **Media Allocation:** selects the media and modalities, from available resources, that will be used to convey the communicative acts.
4. **Ordering:** determines the order in which selected content should be presented during the presentation. The ordering is constrained by the relationships between the communicative acts. Note that multimedia presentations do not necessarily follow the linear structures that written text and speech do.

3.3.3 Design Layer

The transformation of media selected to convey communicative acts into specifications for media objects within an overall presentation layout is a complex process. The production of media objects, and the layout of these objects, are complex tasks that can be broken down into a design task and a realisation task. Also, as previously discussed, the application domain may impose 'standard' layout schema, so there is no justification for assuming that media object production should necessarily precede presentation layout. For these reasons, the standard reference model casts both of these tasks into a Design Layer and a Realisation Layer. The role of the Design Layer is to plan how to convey a communicative act using the allocated media and modalities. This can be broken down into two sub-tasks:

1. **Media Design:** This may include dedicated modules for designing different media and modalities, such as: images, 2D/3D graphics, natural language, animation, video, etc. In some cases the different components may just be required specify the format of existing multimedia content.
2. **Layout Design:** This determines the spatiotemporal arrangement of media objects in the presentation, utilising the available application data and knowledge resources.

There is no particular ordering imposed by the standard reference model for the media and layout design tasks: the media objects desired may constrain the layout decisions that can be made; the layout required for an application may constrain how the media needs to be designed; or they may both impose constraints on each other. The results of the Design Layer are *realisation plans*, which are ordered sets of commands to be executed by the Realisation Layer.

3.3.4 Realisation Layer

The Realisation Layer creates media objects and their layout from their design specifications. As with the Design Layer, the Realisation Layer has two main tasks:

1. **Media Realisation:** This may include dedicated modules for producing different media given the design specifications. Media realisation could be one or more of:
    a. A retrieval task – where the design specifications serve as descriptors that must be matched against available media objects.

b. A formatting task – where sub-elements of available media are selected. For example, a part of an image or a segment of a video.

c. A conversion task – where available media is converted to an appropriate media format. For example, an image may need to be provided in the JPEG format.

d. A generation task – where the design specifications are used to generate a media object matching those specifications. For example, a 2D graphic, animation, or objects moving in a 3D scene (Wark et al., 2009).

2. **Layout Realisation:** This populates the layout specification with the realised instances of the media objects. This task is heavily influenced by the display environment used for the presentation.

The output of the Realisation Layer includes all of the information required to execute the presentation.

3.3.5 Presentation Layer

The Presentation Layer renders the output of the Realisation Layer so that it can be perceived by the user. It coordinates rendering of the various media objects within the display environment, and manages the execution of the presentation in response to user input from the Control Layer, taking into account any resource limitations.

3.3.6 Knowledge Server

The Knowledge Server element of the standard reference model represents those functions that provide knowledge to the different layers. This can be conceptualised as providing four types of expertise:

- **Application Expert:** provides the IMMP system with application-specific knowledge, including:
  - Interface with the application systems
  - Convert information/data into appropriate formats for the IMMP & application
  - Process and make accessible the pool of information from which content can be selected
  - Characterise the incoming information/data so it can be reasoned about by the IMMP system.
- **Context Expert:** This maintains the coherence of the presentation, and is responsible for the resolution of context-dependent references. It has two main tasks:
  - Maintains a representation of what has been generated so far, and the mapping between the media objects and the underlying semantics.
  - Maintains a representation of what has been presented to the user so far, and a representation of the way the user has interacted with it.

- **User Expert:** This maintains a model of the user, which can include representations of:
    - A user's goals and plans (which may be based on their role and needs)
    - A user's physical and mental abilities
    - A user's attitudes and preferences
    - A user's knowledge and beliefs
- **Design Expert:** This complements the other elements of the Knowledge Server by providing all other knowledge which is relevant for decision making by the IMMP system, or should be modelled as a shared resources as it will be accessed by multiple layers. It may include:
    - Models of when media/modalities are appropriate (as per §3.2.4)
    - Design constraints
    - Device models – a partial model of the computational environment and input/output devices.

The Knowledge Server also needs to be able to draw knowledge from external resources, and provide knowledge to other systems to allow the IMMP system to integrate with any other relevant systems.

3.4 Content Generation

As discussed above, a number of issues arise with multimedia presentations that do not occur with text generation systems:

- How can we maintain coherence of a presentation when the content is realised through different modalities?
- How can we make use of images, graphics, animations, etc.? Do they have a consistent internal structure that can be expressed in terms of rhetorical relationships?
- Can we use a common representation to express both textual and graphical communicative acts?

Most multimedia presentation systems have adopted an approach based on the techniques used for text generation, that uses a hierarchy to structure and organise multimedia content, and applied rhetorical relationships to maintain coherence between the elements. Some of the complexity required for an automated IMMP system is captured in the considerations for effective use of multimedia content, and the standard reference architecture discussed above.

The goal of the work being done at DSTO is to apply these IMMP techniques to improve situational awareness for our military clients. While a fully automated IMMP system is desirable, as it will reduce manpower requirements, a semi-automated system using IMMP techniques would also be of value. Our current focus is to provide a multimedia

presentation capability that can be integrated with DSTO's Virtual Advisers to (eventually) provide an automated news service. Within this context, the following capabilities provide a progressively more capable system:

1. A system to allow a user to author and re-use a multimedia presentation to suit a particular audience, time constraints, and display environment.
2. A system to automatically assemble human-authored multimedia content to generate a multimedia presentation on a particular topic, to suit a particular audience, time constraints, and display environment.
3. A system to automatically generate and/or assemble multimedia content to present information on a particular topic, to suit a particular audience, time constraints, and display environment.

The development of a capability to facilitate human-authoring and re-use of multimedia presentations for the Virtual Adviser system has been identified as a pressing need. The current approach using marked up text to specify a presentation suitable for a Virtual Adviser is time consuming, error prone, and requires specialised knowledge. One of the early goals of our work is to provide a simple graphical authoring capability that can assist with the creation of a multimedia presentation using the guidelines discussed in §3.2 to be effective, structured using rhetorical relations as discussed in §3.1.2 to maintain coherence under different presentation constraints, and implemented in a way consistent with the standard reference architecture discussed in §3.3 so as to be compatible with an automated IMMP capability. Initially, this will aim to produce a portable multimedia document that describes the media to be used, and the rhetorical relations between the multimedia elements, without specifying the presentation environment. This will rely on the expertise of the human author to choose and/or create the appropriate multimedia content and assemble it in an effective way to convey the author's communicative goal. An IMMP system would then select the appropriate parts of this presentation, lay it out to suit the presentation environment, and control the presentation to an audience.

3.4.1 IMMP Graphical Editor

A web-based editor for multimedia presentations is planned to allow an author to assemble and appropriately tag content in the presentation so it can be processed by an IMMP system. As discussed above, the editor is not intended to dictate the layout of multimedia content within the presentation, but to allow the structure and rhetorical relationships for the multimedia content to be specified so that an IMMP system can present it. However, it is expected that a preview capability would be useful when authoring a multimedia presentation, and so some capability to specify one or more layouts, even if only for preview purposes, would be useful.

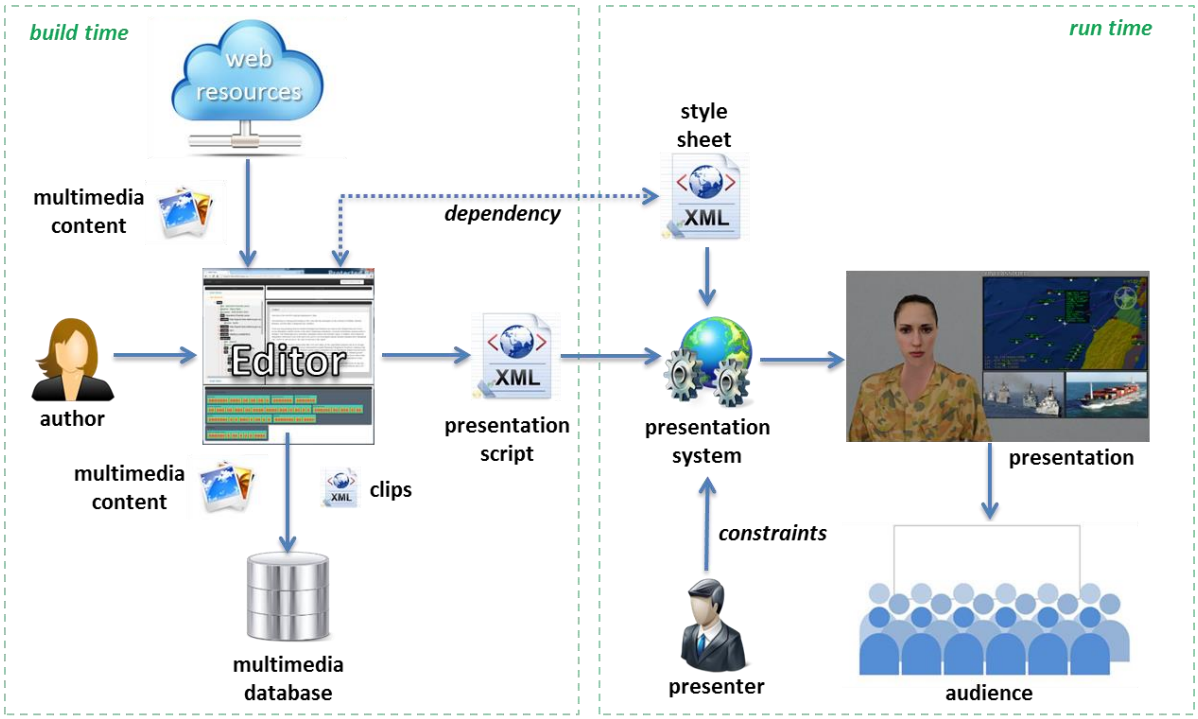


Figure 11: IMMP workflow for proposed editor for authoring IMMP content.

The workflow for the proposed editor is, at least initially:

1. The author assembles text and multimedia content for a multimedia presentation described as an IMMP script.
    a. Text is entered directly by the author, or imported from external sources
    b. Images, graphics, videos, etc. is either imported from third party sources such as the internet, or created in third-party tools and saved as a web resource (e.g. a Wiki)
    c. URIs are used to reference multimedia content (other than text) in the IMMP script.
2. The editor provides templates or hints for the structure of the multimedia content within the presentation. Depending on the user preferences, these could be mandatory or suggestions.
3. Context-specific metadata for the multimedia content used is optionally saved in a database to allow for reuse of generated or discovered content. This could refer to individual multimedia elements, or so-called *clips* of multimedia content assembled to convey a particular concept. In the latter case, the database would save both the metadata and the textual content of the clip and its embedded multimedia references. The editor allows this content to be retrieved using context-specific search parameters so it can be reused in other presentations if desired.
4. The editor allows rhetorical relations to be assigned between multimedia elements in the presentation. A default set of generic relationships is imposed if none is specified.

5. Layout templates, or *styles*, can be specified and saved to support preview of multimedia content during authoring. The style refers to abstract layout design components – it is up to the IMMP system to realise these.
6. The completed presentation can be saved as an IMMP script in an XML format.
7. The IMMP script, along with an optional style sheet specifying the desired layout elements, provides the input to an IMMP system that will select the appropriate content given presentation constraints, and render it as a multimedia presentation.

4. Rhetorical Structure Theory

Rhetorical Structure Theory (RST) was originally developed to support computer generation of text[16] (Mann and Thompson, 1988, Taboada and Mann, 2006), but it is now widely used in linguistics independently from text generation.

RST provides a framework for ensuring coherence of a discourse, and has been generalised to multimedia presentations, by ensuring that every part of a text or multimedia presentation has an evident role described by an RST relation.

4.1 RST Structure

The most frequent structural pattern in a discourse or multimedia presentation is that one discourse element has a specific role relative to another. This is represented as a *nucleus*, and a *satellite* that has a relationship to the nucleus described by a rhetorical relation. A nucleus element may have more than one satellite element, and if a rhetorical relation does not have a particular element which is more central than the other, it is called a *multinuclear* relation. Some simple examples of the structure are shown in Figure 12.

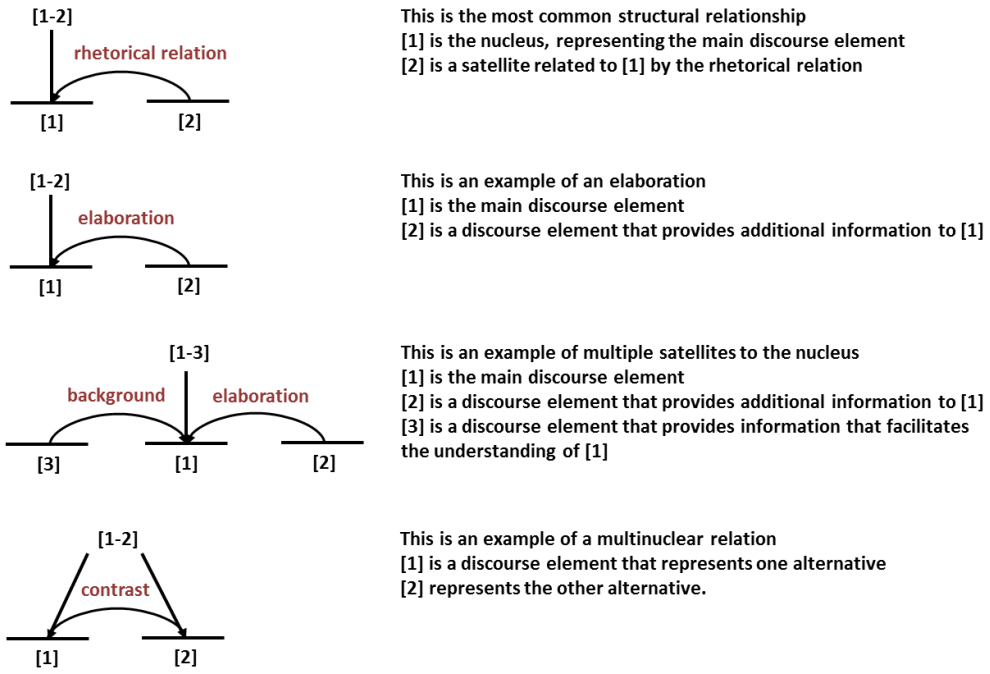


Figure 12: Illustration of structure of common RST relations from (Colineau and Paris, 2003). The red text represents the relation, and arrows point towards the nucleus.

[16] For Mann's account of the origins of RST, see http://www.sfu.ca/rst

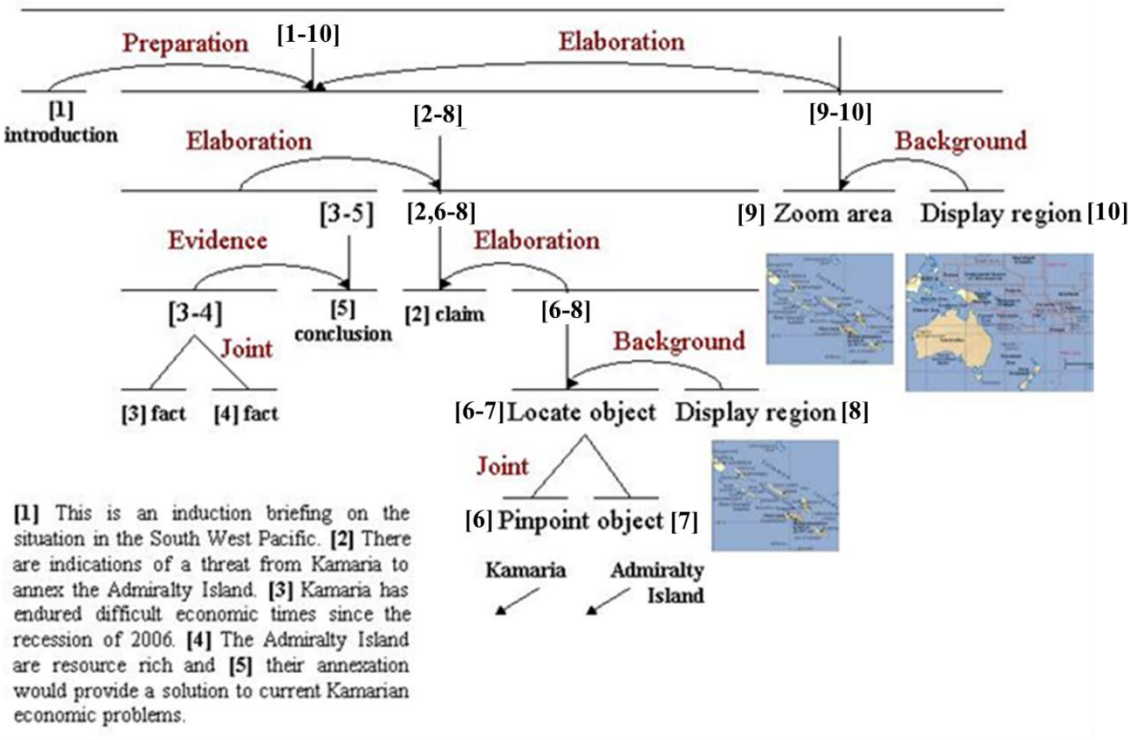


Figure 13: Example of RST applied to a multimedia Induction Briefing developed for FOCAL using a fictitious training scenario, from (Colineau and Paris, 2003, Paris et al., 2004) adapted to label all multimedia elements used.

Figure 13 shows how these relations can be applied to a multimedia presentation (Colineau and Paris, 2003, Paris et al., 2004), in this case an Induction Briefing taken from a fictitious training scenario used in FOCAL (Wark et al., 2004). We can see in this example that the rhetorical relations are organised as a hierarchy.

At the top most level, the presentation is organised into three elements:

- The nucleus, which is a complex element composed of text elements [2-5] plus additional illustrations [6-8]
- Two satellite elements:
    - One of the satellites ([1]) is linked to the nucleus by the RST relation called *preparation*, which indicates that it presents information that introduces the content contained in the nucleus
    - The other satellite is a complex element composed of two illustrations [9-10] of the region that represents an *elaboration* of the nucleus providing additional detail. This satellite can itself be decomposed into a nucleus [9], presented as a zoomed view of the area of interest, and a satellite [10] linked to it by the RST relation called *background*, which indicates that it facilitates understanding of the nucleus – in this case by showing a larger scale map situating the area of interest with respect to Australia.

The nucleus of the presentation is itself composed of three elements:

- A nucleus containing the main claim ([2]) of the Induction Briefing
- Two complex satellite elements:
    - One satellite ([3-5]) is a complex element that is an elaboration of the main claim, which can be similarly decomposed.
    - The other satellite ([6-8]) is also a complex element that contains graphics that provides additional elaboration, which can also be similarly decomposed.

In this way, all of the multimedia elements of this example can be linked to another via the RST relations, providing a coherent presentation.

4.2 RST Relations

The RST relations can be classified into nucleus-satellite and multinuclear relations as discussed earlier. They can also be classified according to the intended effect on the audience:

- **Presentational relations:** are those where the intended effect of the satellite is to induce an attitude in the audience about the nucleus, such as a desire to act, a positive (or negative) regard of, a belief in, an acceptance of.
- **Subject Matter relations:** are those where the intended effect of the satellite is to inform the audience.

In all there are currently some 32 RST relations that have been defined. A point to note about the names assigned to the relations used in Rhetorical Structure Theory is that they do not necessarily accurately reflect the intent of the relations – some inconsistencies have arisen because there is only a limited pool of names available[17].

4.2.1 Presentational Relations

Table 1: Presentational relations used by Rhetorical Structure Theory

Relation	Nucleus	Satellite
Antithesis	Ideas favoured by the author	Ideas disfavoured by the author
Background	Content whose understanding is being facilitated	Content intended to facilitate understanding
Concession	Situation affirmed by the author	Situation which is apparently inconsistent but also affirmed by the author
Enablement	An action	Information intended to aid the audience in performing an action
Evidence	A claim	Information intended to increase the audience's belief in the claim

[17] See http://www.sfu.ca/rst

Justify	Content	Information supporting the author's right to express the content
Motivation	An action	Information intended to increase the audience's desire to perform an action
Preparation	Content to be presented	Content which prepares the audience to expect and interpret the content to be presented.
Restatement	A situation	A re-expression of the situation intended to increase the audience's awareness of the situation
Summary	Content	A (short) summary of the content intended to increase the audience's understanding of the content

4.2.2 Subject Matter Relations

Table 2: Subject Matter Relations used by Rhetorical Structure Theory

Relation	**Nucleus**	**Satellite**
Circumstance	Content expressing events or ideas to be interpreted	The context in which the content is to be interpreted.
Condition	Action or situation whose occurrence results from another	Condition resulting in the action or situation
Elaboration	Basic (core) information	Additional information
Evaluation	A situation	Assessment of nucleus
Interpretation	A situation	An interpretation of a situation
Means	A situation	A method or instrument that makes realisation of the situation more likely
Non-volitional Cause	A situation	Another situation which causes the other, but not by deliberation action
Non-volitional Result	A situation	Another situation caused by the other, but not by deliberate action
Otherwise	Action or situation whose occurrence results from non-occurrence of another	Condition resulting in another action or situation
Purpose	An intended situation	The intent behind the situation
Solutionhood	A situation or method supporting full or partial satisfaction of the need	A question, request, problem, or other expressed need
Unconditional	Action or situation	Another action or situation which the nucleus does not depend on
Unless	Action or situation	Another action or situation which will prevent the nucleus from occurring
Volitional Cause	A situation	Another situation which causes the other, by deliberate action
Volitional Result	A situation	Another situation caused by the other, by deliberate action

4.2.3 Multinuclear Relations

Table 3: Multinuclear Relations used by Rhetorical Structure Theory

Relation	Element	Other Element
Conjunction	Part of a unit	Another part of a unit that plays a comparable role
Contrast	One alternative	The other alternative
Disjunction	One alternative	Another alternative
Joint	Unconstrained	Unconstrained
List	An item	Another item
Multinuclear Restatement	An item	A restatement of comparable importance
Sequence	An item	The next item

4.3 A Graphical Representation of RST

For the purposes of our initial work we can represent the hierarchical structure of a presentation as a directed graph. In this representation, the nodes represent combinations of one or more multimedia elements, and the edges represent the relationships between them. In this case, the RST relations can be assigned to the edges between a satellite and its nucleus. In addition, we assign a 'nucleus' relation to the edge between a composite node and its constituent nodes (i.e. the vertical lines in Figure 13).

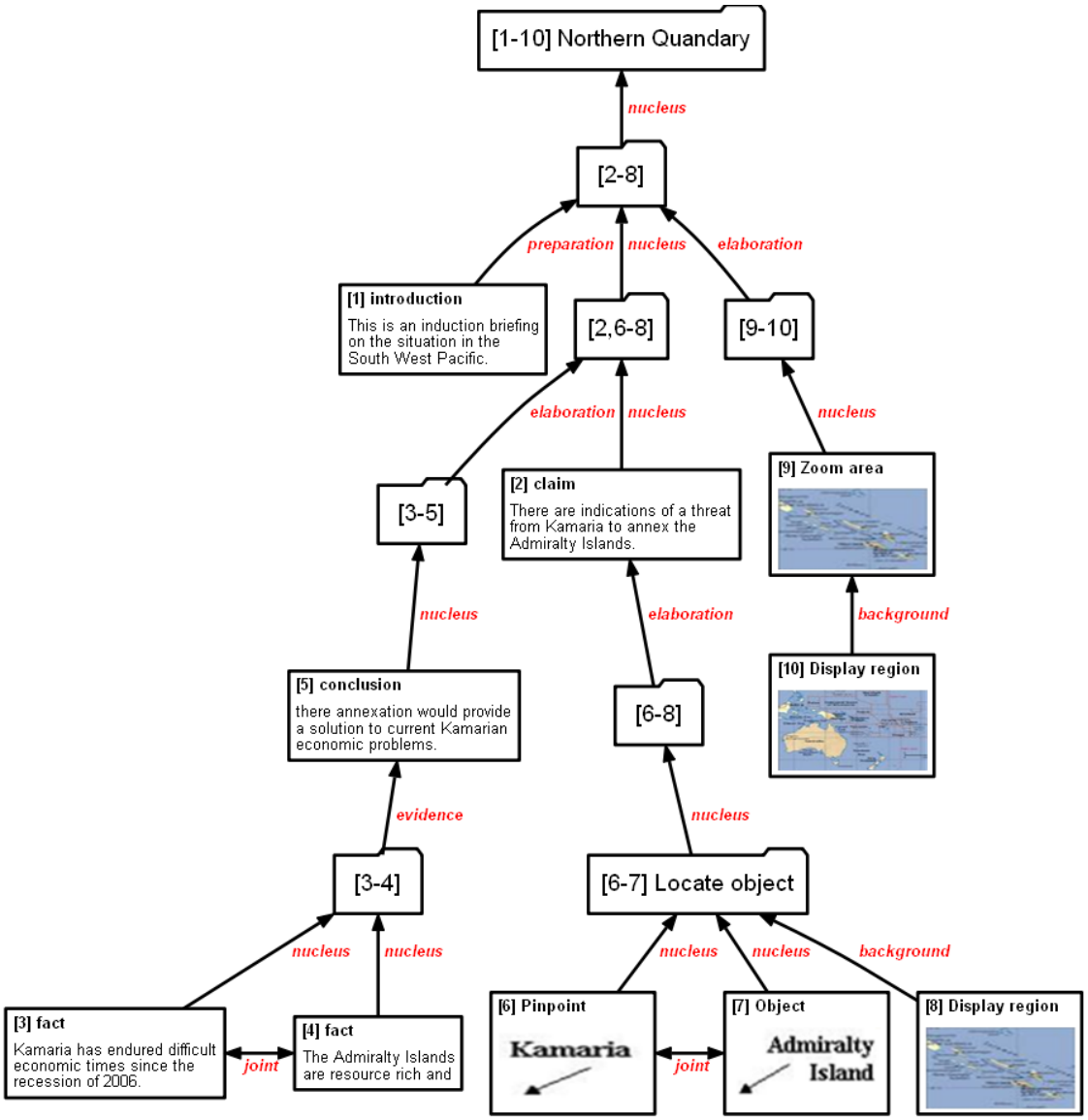


Figure 14: Graphical representation of the Northern Quandary Induction Brief, mapping the RST relations to the edges of the graph, and with the inclusion of a 'nucleus' relation that indicates that the source node forms the nucleus of a larger discourse element.

This representation can be used to apply different selection strategies to the graph that maintain overall coherence of the presentation, by maintaining connectivity to the root node. From inspection of the example shown in Figure 14, we can make some observations regarding content selection:

1. A satellite for a composite node does not maintain coherence with the overall presentation without the nucleus of that composite node.
2. The nucleus of a composite node is required to maintain coherence within that node.

3. All nuclei in a multinuclear relationship are required to maintain coherence within that composite node.

Thus, we can see that *for this purpose,* assuming that *edges with the nucleus and multinuclear relationships are never broken,* we can simplify the graph, by:

1. Replacing each composite node with its nucleus, which inherits the satellites of the composite node.
2. Replacing multinuclear relationships with a single nucleus (this relies on our assumption above that multinuclear relationships are never broken).
3. Retaining the root node of the graph for reference purposes. The nucleus relationship is now only retained at this level in the graph, and indicates that this node is the nucleus of an RST relationship but not the satellite of any other node in the graph.

This produces a tree whose branches can be pruned (given the constraints above) while retaining overall coherence – i.e. rhetorical relations exist between all the elements. An example is shown in Figure 15.

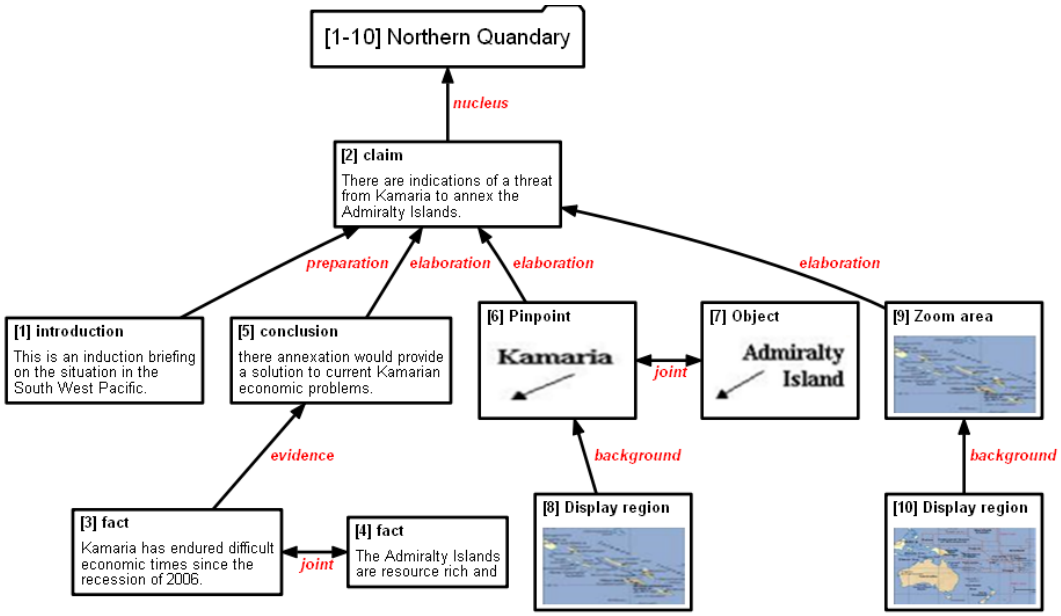


Figure 15: Simplified representation for Northern Quandary Induction Brief provides a functional map for content selection that maintains coherence.

This graphical representation of a presentation could provide a useful template for synthesis of a multimedia presentation that will retain coherence as content is pruned. One of the aims of our work was to explore the feasibility of this approach.

4.4 Simplified Rhetorical Relations

For our purposes the complete set of 32 RST relations were not considered necessary (at least initially), so a simplified set was chosen to permit evaluation of the feasibility of our approach. Additional relations were added to handle special cases arising from the way we have structured our multimedia presentations.

There were 6 primary rhetorical relations chosen:

- **Preparation:** as per the RST relation, this establishes the narrative context for the content in the nucleus.
- **Elaboration:** this is a generalisation of many RST relations (including *elaboration*), providing more information about the nucleus and usually presented after it.
- **Joint:** this is a generalisation of all of the RST multinuclear relations. Its primary purpose in our work is to indicate that several multimedia elements need to be presented together. For our work we assumed implicit sequencing of content based on the order in which it appears in the document. The multinuclear *sequence* relationship is subsumed by the *joint* relation in this case.
- **Background:** as per the RST relation, this content facilitates understanding of the nucleus by providing some situational context, and is usually presented either before the nucleus or alongside it.
- **Conclusion:** this was introduced as a narrative construct to finalise a particular discourse. It is functionally similar to the *preparation* for a nucleus.
- **Summary:** as per the RST relation, this provides a short restatement of the nucleus to provide reinforcement.

An additional rhetorical relation was added to support our implementation:

- **Initialisation:** this is a special case of *preparation*, intended to ensure that all multimedia content is appropriately initialised for the subsequent discourse. Unlike the *preparation,* which may or may not be included with its nucleus, the *initialisation* must *always* be presented before its nucleus. This provides a convenient way of ensuring that multimedia display channels are appropriately initialised within the presentation.

As discussed earlier, we also include a subsumption relation:

- **Nucleus:** within the decomposition of a composite discourse structure, there will always be one element that is the nucleus of an RST relationship but not the satellite of any other element within that structure. We designate this element as the *nucleus* of the composite discourse structure. In our computational implementation this relationship serves to link the node representing the composite discourse structure with the nodes contained within that structure, as illustrated in Figure 15. While this differs from the usual notion of the nucleus in RST in a nucleus-satellite relationship, it is not inconsistent with it.

These relations are used to tag the multimedia content within a multimedia presentation, to facilitate content selection. If this approach proves useful, generalisations such as elaboration and joint could be expanded to allow fine-tuned selection of content, and integration with automated IMMP approaches.

5. An XML Format for Multimedia Presentations

The multimedia presentation generated by the proposed graphical editor will be saved as an XML document that includes the textual components, references (URI) to other media, and the rhetorical relationships between them, to provide portability and rendering with different layouts. The requirements identified for the presentation format were:

- Contain information about the content of a presentation without requiring specification of the final presentation layout
- Allow multimedia elements to be collected in a discourse element that represents a particular concept
- Allow tagging of discourse elements with rhetorical relations to allow a coherent presentation to be produced under different presentation constraints
- Allow a hierarchy of discourse elements to be constructed, linked by rhetorical relations.
- Allow collections of discourse elements to be tagged with semantic information so they can be re-used in other presentations

A number of approaches were considered, including some existing multimedia standards, but none were found to have the features required. For this reason, a bespoke XML format was developed and subsequently trialled.

Much of the rationale behind the design of this format came from earlier work done to provide a multimedia presentation system for FOCAL (Wark et al., 2004). In this work, a scenario document based on the ADF's Northern Quandary training scenario was converted into an XML format based on the structure of the document (without rhetorical relations). This grouped multimedia content together so it could be presented together as a unit, and allowed the ATTITUDE (LAMBERT, 1999) multi-agent system to control the playback of the presentation. In this case, a dialogue management system was integrated with the presentation system so that the user could control the playback and query the system to retrieve content (Estival et al., 2003).

5.1 SMIL

The Synchronised Multimedia Integration Language (SMIL)[18] (version 3) is a W3C standard developed to enable simple authoring of interactive audio-visual presentations. It was considered as a candidate output format for our IMMP editor, but was found to be unsuitable because:

[18] See http://www.w3.org/AudioVideo/

- SMIL is aimed at an implicitly single-screen, 2D presentation environment, such as a web browser. As such, the layout specifications used for the multimedia content is restricted to a 2D coordinate system.
- The layout used for multimedia content is embedded within the SMIL document. This does not support reuse of the content with different presentation environments.
- SMIL does not support the inclusion of tags for the rhetorical relations associated with multimedia content.

Like other multimedia standards such as MHEG[19], SMIL is more suited as a specification for a multimedia presentation after the realisation stage in the standard reference model.

5.2 SMPL

The 'Simple Multimedia Presentation Language' (SMPL) was an early attempt to produce an XML format that embeds rhetorical relations within a human authored document to simplify content generation and re-use for the Virtual Adviser system. SMPL was structured around a small set of rhetorical relations that applied to a fixed hierarchy of discourse elements:

- **Nucleus:** a collection of multimedia elements that represents the core communicative act. Presentation of this is prioritised over other content. It was typically provided as an utterance for the Virtual Adviser.
- **Elaboration:** a collection of multimedia elements that provides additional information about the nucleus.
- **Caption:** a textual cue to the content conveyed in the nucleus. This represents preparation for the nucleus, and was typically provided as a caption to the Virtual Adviser.
- **Topic:** a collection of multimedia elements that provides a contextual cue for the content presented by one or more discourse elements (nuclei and their satellites) with a common topic, grouped into a collection of discourse elements dubbed a *clip*.
- **Background**: a collection of multimedia elements that provides common background information for a collection of clips.

SMPL was found to provide a useful format for helping the author structure the presentation, and dynamically determining the appropriate layout of content in different presentation environments. However, its rigid structure allowed only limited selection of content to satisfy presentation constraints, and it was not considered suitable for a more adaptive capability.

[19] MHEG is an object-based multimedia standard developed primarily for interactive news services. See http://www.mheg.org/users/mheg/

5.3 The Proposed Format

The format adopted, at least initially, for the multimedia presentations produced by the editor is based on the SMPL format discussed above, but without its rigid rhetorical relation structure. A simple three-level structural hierarchy is adopted for a presentation:

1. Multimedia content is organised into *segments*.
2. Collections of segments are contained within a re-usable multimedia *clip*, and rhetorical relationships between the segments within the clip determine the narrative structure of the clip. Each clip is intended to stand alone, and to convey a particular communicative (sub) goal.
3. Clips are combined into *sequences* to achieve the overall communicative goal of a presentation. The overall narrative structure of the presentation is determined by the rhetorical relationships between clips.

This structure is illustrated in Figure 16.

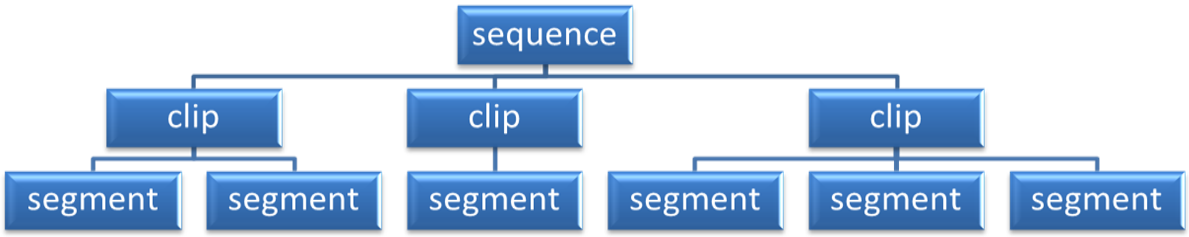


Figure 16: Three-level structural hierarchy adopted for multimedia presentations allows aggregation of multimedia segments into re-usable clips that can be combined to form a presentation described as a sequence of clips.

This structure does not, however, limit the complexity of the hierarchy of discourse elements that can be contained within the multimedia document, as an arbitrarily complex narrative structure can be imposed via the rhetorical relations between the clips, and between the segments within them.

5.3.1 IMMP Structure

There are two important dimensions defined in the XML format: the multimedia content within the presentation and the rhetorical relations between them; and the way it is to be rendered. While in a purely automated IMMP system the latter would be determined automatically, for a scheme intended for human authoring it was considered important to allow the human to provide some level of design input, as well as simplifying the task of previewing the presentations created. In the format chosen, the author specifies an *abstract* rendering *channel* for the each multimedia element. How this information is ultimately interpreted by an IMMP system determines how this influences the presentation design and realisation.

5.3.1.1 Content

The multimedia content within the XML document is structured using the XML elements below:

- **content:** represents an single piece of multimedia information in the presentation. For example, this could be:
  - a sentence, paragraph, or bullet point
  - an image, 2D graphic, or 3D scene
  - a video, 2D or 3D animation
  - a formal specification for multimedia content (e.g. X3D[20] specifies a 3D animation)

  Each content element uses an abstract *channel* (see 5.3.1.2) to specify how it is to be rendered.
- **segment:** represents a temporally coordinated set of multimedia *content*. Segments allow multimedia content to be combined where each element by itself may not add value to a presentation. For example, this could be a text caption with an image, or narration of a video. While each segment does contain an implicit narrative structure, it was not considered necessary to explicitly specify it at this fine-grained level as it imposes an additional burden on the author. Segments thus form an atomic multimedia element within our presentation system.
- **clip:** represents an aggregation of *segments* that could be used 'stand-alone' to convey a concept (or related concepts). Clips could potentially be tagged, retrieved and assembled as multimedia content in their own right. The narrative structure of a clip is determined by rhetorical relationships between the segments within the clip.
- **sequence:** represents a collection of *clips* that can be used to make up a multimedia presentation. The overall narrative structure of the presentation is determined by the rhetorical relationships between the clips.
- **script:** represents one or more *sequences*, possibly on related topics. In most cases a script may well include only a single sequence, but our experience has been that in some cases it is useful to be able to bundle presentations on a common theme together. For example, each sequence could represent a particular act in a play, or phase of a demonstration.

5.3.1.2 Styles

The multimedia content within the XML document is notionally assigned to the abstract rendering constructs described below, represented as attributes of the associated content elements. The IMMP system determines how these are interpreted.

[20] See http://www.web3d.org

- **channel:** represents an abstract rendering mechanism for the associated *content* element. The intended rendering mechanism will often determine the content generated by the author. For example:
    - text could be rendered into an 'utterance' or 'caption' channel. As a caption only a word or two might be used, while as an utterance a complete sentence or paragraph is more likely.
    - An image could be rendered into a 'monitor' window (à la a TV news service) or as a background. More detailed information is likely to be included in the former, while more contextual information is appropriate for the latter.
- **style:** represents an abstract formatting or transition effect to be applied to the associated *content* element. For example, this could be used to indicate:
    - a fade-in, fade-out, swipe for an image or graphic
    - the font, colour, and size of text
    - the facial expression, gestures, or mood to be used by a Virtual Adviser when delivering an utterance.
- **layout:** represents an abstract grouping and arrangement of *channels* and *styles* associated with a *clip* element. A default *layout* can be associated with a *sequence*, which is used for any clip elements within the sequence that do not have an associated layout.
- **stylesheet:** represents the realisation of the abstract *channels*, *styles*, and *layouts* used in a presentation document that determines the design of the presentation. This may be an artefact, such as an XML document, or a presentation 'mode' applied by the IMMP system.

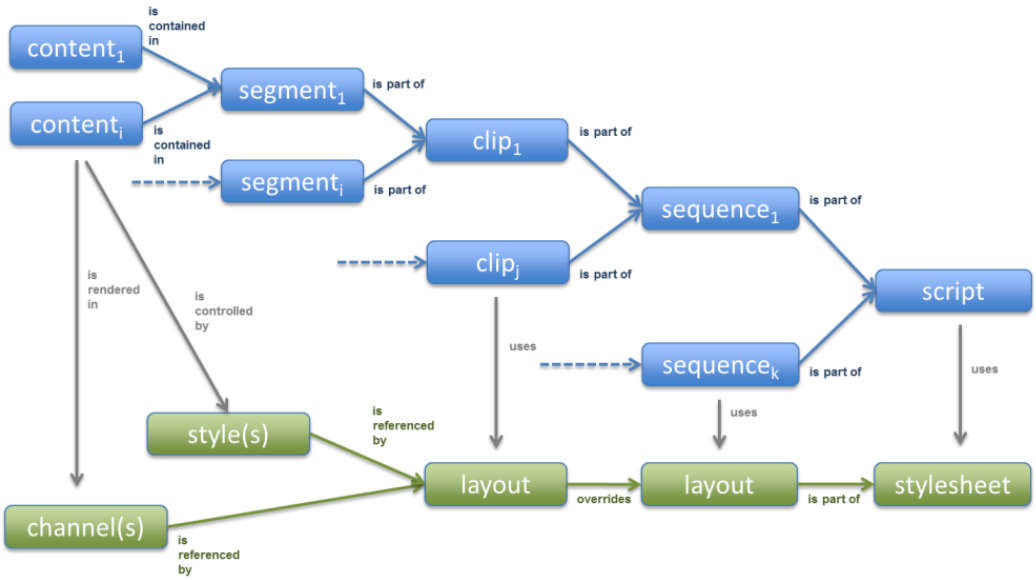


Figure 17: Structural relationships between XML elements used in the multimedia presentation document.

5.3.2 Topics

In order to potentially retrieve saved clips for re-use and assembly as part of another presentation, one or more XML *topic* elements can be added to a clip. A topic is represented by an abstract *name* and *ontology*. This also allows an IMMP system to only extract that multimedia content from a presentation related to a particular topic.

5.3.3 Rhetorical Relations

Rhetorical relations are represented as XML elements within the multimedia document[21]:

1. Rhetorical relations are assigned to the *segments* within a *clip*. In this case, one segment is assigned as the implicit *nucleus* (see §4.4) of the clip.
2. Rhetorical relations are assigned to the *clips* within a *sequence*. In this case, topic and ontology attributes can also be associated with the rhetorical relation, as at this level the rhetorical relationships between clips may be different for different topics. One clip is assigned as the implicit *nucleus* (see §4.4) for a particular topic within a sequence.

As discussed in §4.4, there are 7 rhetorical relations currently supported within the multimedia document:

- **Preparation**
- **Initialisation**[22]
- **Elaboration**
- **Joint**
- **Background**
- **Conclusion**
- **Summary**

The interpretation of these relations is handled purely by the IMMP system, so this set can be easily extended without requiring changes to the document schema.

The rhetorical relations link a satellite[23] element with its nominated nucleus, which in turn links via a rhetorical relation to *its* nucleus, etc. This scheme provides a linked list of relationships between the elements (segments or clips) within a clip or sequence (respectively). This provides sufficient scope for an author to synthesise a complex narrative structure.

---

[21] Initially, rhetorical relations were also assigned to *content* elements within the document, but after some trial of this approach it was determined to add significant overhead but no significant value. If this level of granularity is required it can be obtained, in most cases, by managing the content contained within a *segment*.

[22] As discussed in §4.4, we introduced this relation as a special type of preparation.

[23] In our implementation we consider multinuclear relations to be represented by a single nucleus and a set of special satellites that are always associated with the nucleus.

6. The Example Scenario

The scenario chosen for this initial evaluation of our approach was the Blueland Intelligence Organisation (BIO) intelligence briefing used in DSTO's Integrator demonstration. This is loosely based on the fictitious 'Military Strikes in Atlantis' scenario (Blanchette, 2005) used by The Technical Cooperation Program's (TTCP) C3I Technical Panel on Information Fusion, and was chosen because it provided a militarily feasible scenario for usage of the IMMP system being developed. In this scenario, different length briefings were required depending on the knowledge of the audience about the situation.

6.1 Military Strikes in Atlantis

This scenario is based in the fictitious continent of Atlantis, located in the North Atlantic Ocean to the West of continental Europe. Atlantis is composed of 6 countries: Blueland, Redland, Brownland, Orangeland, Whiteland and Greyland. For the purposes of this scenario, the locations of Redland and Blueland have been reversed from that found in the original TTCP scenario. There has been a long-running dispute between Redland and Blueland over sovereignty of the Camrien Peninsula, and Redland has launched a surprise invasion of Blueland controlled territory. The UNSC has issued a resolution requesting that Redland withdraw from the Camrien Peninsula, but it does not intend to comply. However, Redland needs to resupply its forces in the Camrien Peninsula, and it is likely that it will source munitions from outside of Atlantis. The Blueland Defence Force has blockaded the maritime routes to the Camrien Peninsula, and are using its intelligence and surveillance assets to identify, and board, any vessels suspected of carrying munitions. Blueland surveillance has detected a cargo vessel, suspected of carrying munitions, escorted by a Redland FFG, approaching the Camrien Peninsula.

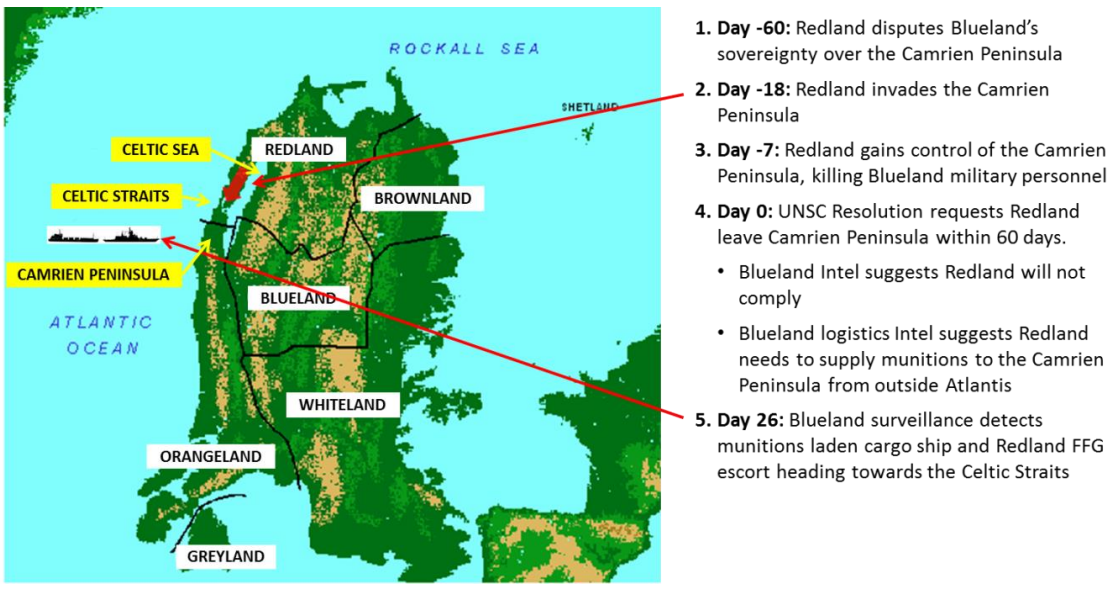


Figure 18: Vignette of the Military Strikes in Atlantis scenario used as a test case for evaluation of the IMMP system.

6.2 BIO Intelligence Update

The Intelligence Update provided to the Blueland Intelligence Organisation is used as the test case for our IMMP work. In the DSTO Integrator demonstration, this was provided by a Virtual Adviser, using hand coded THML. For our initial work, in order to evaluate the feasibility of our approach prior to development of an IMMP editor, this briefing was manually transcribed into our XML format, assigning different multimedia elements to media channels aligned with those used in the demonstration, and organised into segments corresponding the synchronised multimedia content used in the demonstration.

In this presentation, 5 multimedia channels were utilised:

- **narration:** utterances for the Virtual Adviser
- **caption:** text to be used as a caption for the Virtual Adviser
- **icon:** a graphic associated with the caption
- **monitor:** a virtual 'video' screen showing an image, video, or graphic
- **vb:** a script describing a 3D scene or animation in the Virtual Battlespace software (Wark et al., 2009)

While this assignment of channels represents how this content was used in the source presentation, in the XML document these are abstract representations only, and how they are interpreted to produce the final presentation is determined by the IMMP system.

Segments were grouped into three main discourse elements that dealt with different concepts, and these formed the basis of three clips used in our representation. These clips were:

- **Introduction**: describing the nature of the presentation and the equipment being used
- **Background**: describing the situation that has developed in Atlantis
- **Update**: describing the current situation and the tasking assigned to BIO

Rhetorical relations were assigned between the clips, based on three nominal topics that represent different audience perspectives:

- **Atlantis:** focussed on the perspective of the situation in Atlantis
- **BIO:** focussed on the perspective of BIO
- **BIS:** focussed on the perspective of the equipment being used by BIO

Each of these topics applies to a different subset of the clips in the presentation sequence, and each uses different rhetorical relations between the clips. Hence the discourse structure of the presentation is different for each of these topics, and the presentation achieves a different communicative goal. For example: the 'Atlantis' topic spanned all of the clips in the presentation sequence, with the 'Update' as the nucleus of this presentation sequence; but the 'BIS' topic only contains the 'Introduction' clip, with it as the nucleus of this presentation sequence.

Within each clip, the rhetorical relations between the segments of the clip were assigned based on what constituted the core content of the clip (the nucleus), and the relationships of the other segments to this. Because the intent of the clip structure is to provide re-usable multimedia content in a variety of roles, the rhetorical relations within a clip should be independent of the topic. The degree to which the rhetorical relations within a clip could depend on the topic thus determines the granularity applied to the content within a clip. In order to ensure that each clip can be used independently of whatever content has been presented before it, an *initialisation* segment is assigned within each clip[24]. The resulting narrative structure produced for the presentation with the 'Atlantis' topic, is shown in Figure 19. The complete XML document is provided in Appendix C.

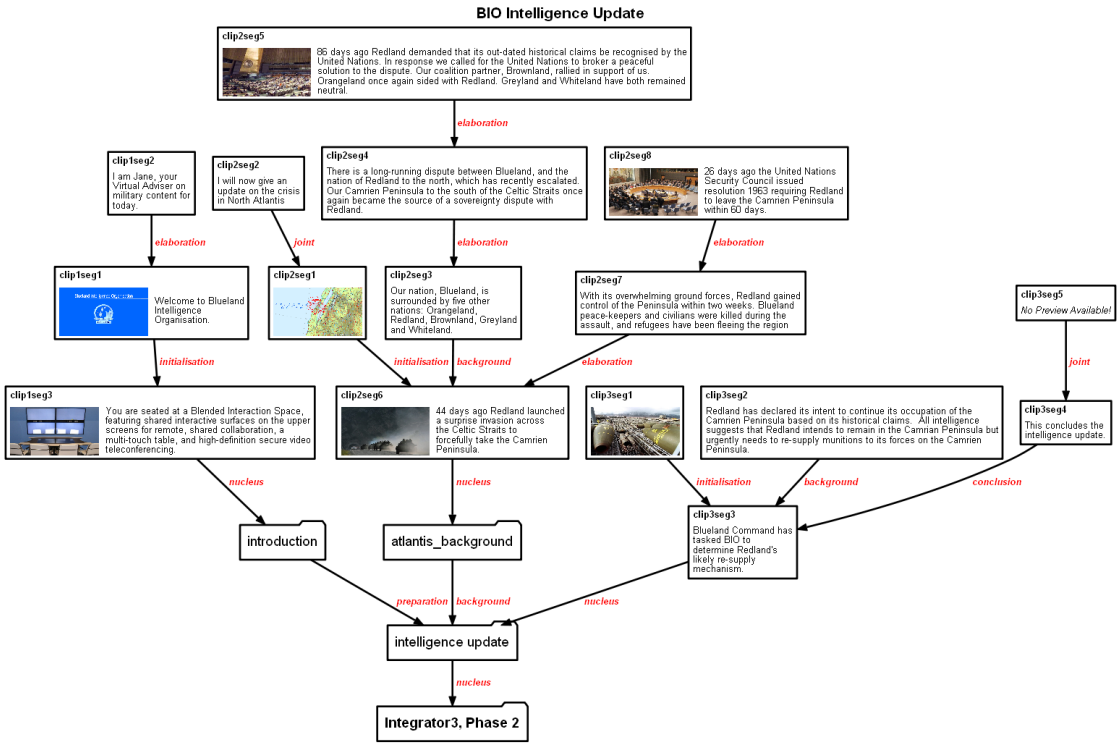


Figure 19: Overall structure of BIO Intelligence Update showing rhetorical relations assigned to multimedia content for the 'Atlantis' topic. For purposes of simplification, only the 'monitor' and 'narration' rendering channels for each segment are shown.

6.3 Scenario Storyboard

The XML document produced can be rendered as a HTML storyboard, showing the multimedia content assigned to the different channels, segments, and clips, as shown in Figure 20.

[24] In the absence of an initialisation segment, the IMMP system would need to decide how to deal with transitions between clips.

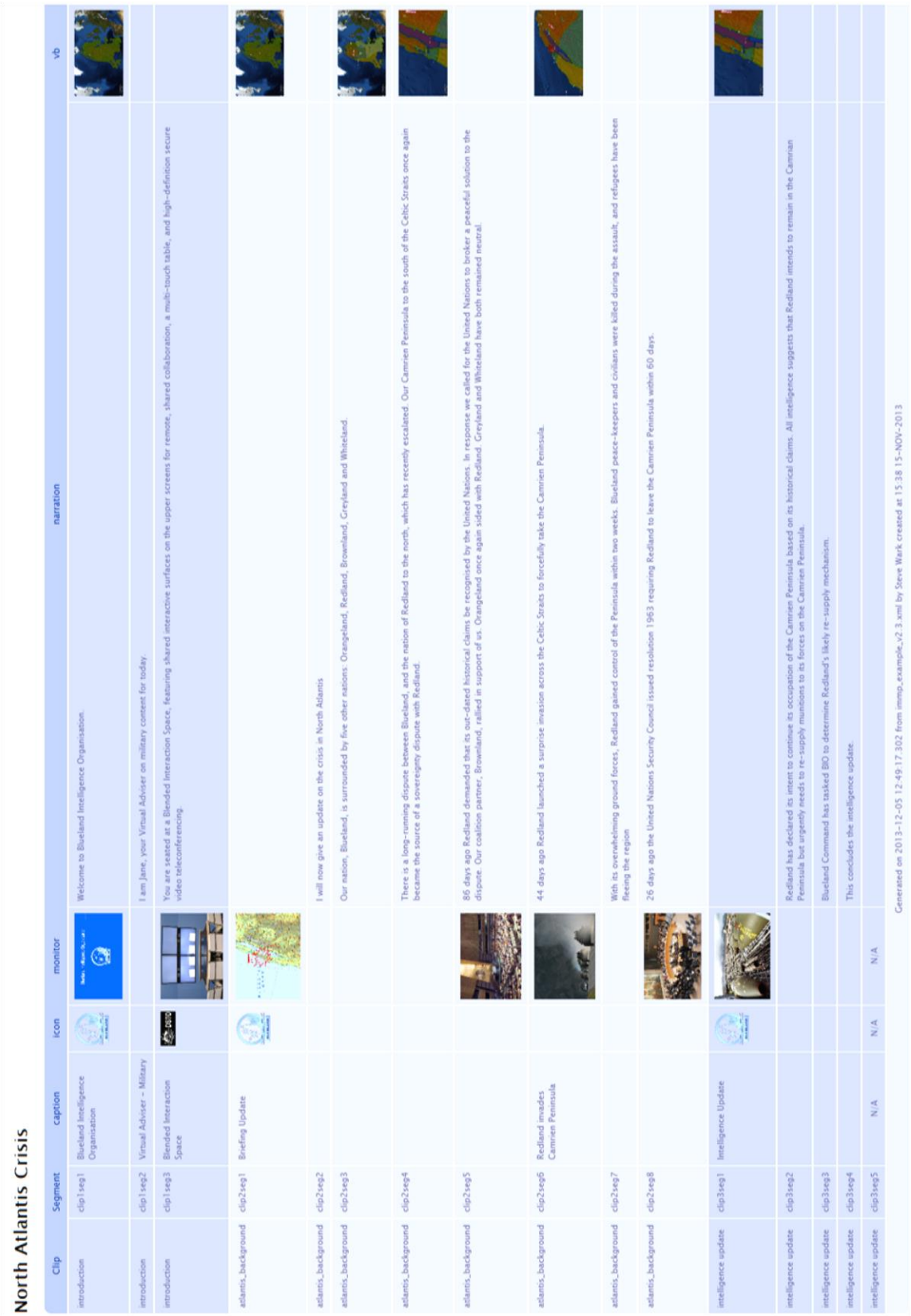


Figure 20: HTML rendering of multimedia document for BIO Intelligence Update.

7. IMMP Content Selection

The example multimedia presentation document discussed in §6 provides a useful test-case to explore the feasibility of various IMMP techniques. Our initial goal was to look at how, given a multimedia presentation in this form with assigned rhetorical relations, content could be selected to provide different presentations constrained by:

1. Topic
2. Prior knowledge of the audience
3. Duration

The effectiveness of the strategies explored was determined by evaluating whether the narrative coherence of the generated presentations was maintained, both in the formal sense and in a subjective sense. The former, which requires that every multimedia segment retains a rhetorical relation to another, is easily established; but the latter requires that the generated presentation still makes 'sense' to an audience, and depends on how well the identified multimedia elements and their rhetorical relations have been defined.

7.1 Selection Approach

The three dimensions to the selection of content from the multimedia presentation can be achieved by:

1. **Topic:** given the XML format used for the multimedia presentation, this is relatively straightforward – only those clips tagged with the specified topic (and ontology) are considered, and the rhetorical relations used are those appropriate for that topic.
2. **Prior Knowledge of the Audience:** the rhetorical relations between multimedia elements determine how they are related to each other. By favouring the selection of content related by particular rhetorical relations over others, the presentation can be slanted towards a particular style that suits different audiences. For example, selecting summary over elaboration may be appropriate for an audience just requiring a review, while selecting background over summary may be appropriate for an audience unfamiliar with the context of the topic presented.
3. **Duration:** The more multimedia elements within a presentation, the longer it will generally take to present. For a time-limited presentation, multimedia elements may need to be pruned to allow the presentation to fit within a nominated duration. The tree structure (and constraints) adopted in our approach should allow branches to be pruned from the graph while maintaining the core structure of the presentation and retaining formal coherence. The strategy applied to prune the tree will determine what rhetorical relations are favoured over others, and the resulting structure of the presentation produced for a nominated duration..

There are a number of approaches that can be applied to determine what content needs to be included and what can be pruned. It was decided to test, at least initially, a simple quantitative approach based on assigning different weights in the range [0,1] to the

rhetorical relations, where 0 indicates it is unimportant to conveying the communicative goal of the presentation, and 1 indicates it is most important to conveying the communicative goal of the presentation. To allow for the presentation structure to influence the selection of content, the weighting applied is also a function of the type of multimedia element (segment or clip).

To evaluate how important any particular clip is to conveying the communicative goal of the presentation, the product of the weights of all of the edges traversed to reach the designated 'nucleus' of the multimedia presentation sequence is calculated (see, for example, Figure 19).

Let:

$\{R\}$	= set of rhetorical relations (plus nucleus)
$\{S(s,c)\}$	= set of segments linking segment s of clip c to clip nucleus
$\{C(c)\}$	= set of clips linking clip c to sequence nucleus
$R_{segment}(s,c)$	= rhetorical relation assigned to edge from segment s of clip c
$R_{clip}(c)$	= rhetorical relation assigned to edge from clip c of sequence
$w_{segment}(R) \in [0,1]$	= weighting of edge from segment with Rhetorical relation $R \in \{R\}$
$w_{clip}(R) \in [0,1]$	= weighting of edge from clip with Rhetorical relation $R \in \{R\}$

and

$W(s,c) \in [0,1]$	= score for segment s of clip c

Then:

$$W(s,c) = \left\{\prod\nolimits_{i\in\{C(c)\}} w_{clip}(R_{clip}(i))\right\} * \prod\nolimits_{j\in\{S(s,c)\}} w_{segment}(R_{segment}(j,c))$$

Obviously, the expectation is that the 'nucleus' of a clip or sequence represents the most important content for conveying the communicative goal, so:

$$w_{clip}(nucleus) = w_{segment}(nucleus) = 1$$

Similarly, there should be no discrimination between content related by the *joint* relationship, so:

$$w_{clip}(joint) = w_{segment}(joint) = 1$$

Finally, within a clip the initialisation relationship is important to allowing the clip to be used independently of the previously presented content, so:

$$w_{segment}(initialisation) = 1$$

Given that each multimedia presentation sequence should contain a 'nucleus' clip, and that each clip should contain a 'nucleus' segment, with these boundary conditions there will always be at least one segment in the multimedia presentation with a score of 1. Given a nominal selection threshold in the range [0,1], segments with a score less than the threshold are deemed less important to the communicative goal than content with a score higher than the selection threshold. This provides a basis by which, given, say, time or

other constraints, content can be included within a presentation. In this scheme, different assignments to the other weightings w_{clip} and $w_{segment}$ will determine what rhetorical relations are deemed to be more important, and the overall structure of the presentation produced. For example, if we consider the special case of the 'nucleus' for each clip, the weighting will be:

$$W(nucleus, c) = \prod_{i \in \{C(c)\}} w_{clip}(R_{clip}(i))$$

while the weighting of each segment of the 'nucleus' clip will be:

$$W(s, nucleus) = \prod_{j \in \{S(s,c)\}} w_{segment}(R_{segment}(j, nucleus))$$

So, if we choose weightings such that:

$$w_{clip}(R) > w_{segment}(R) \ \forall R \in \{R\}$$

Then, where we have a corresponding set of rhetorical relations, we will have:

$$W(nucleus, c) \geq W(s, nucleus) \ \forall c, s$$

So, the segments forming the nuclei of each clip will have a higher weighting than the satellite segments of the clip that forms the 'nucleus' of the presentation (see Figure 21a). The result will be that, given constraints, the presentation will retain the core (nucleus) of each clip in preference to the full content of the core (nucleus) clip. This will tend, as constraints are applied, towards providing an overview of the presentation addressing multiple communicative sub-goals. Conversely, if

$$w_{clip}(R) < w_{segment}(R) \ \forall R \in \{R\}$$

then

$$W(nucleus, c) \leq W(s, nucleus) \ \forall c, s$$

for a corresponding set of rhetorical relations. In this case, the satellite segments of the clip that forms the 'nucleus' of the presentation will have a higher weighting than the nuclei of each clip (see Figure 21b). The result will be that the presentation will retain the full content of its core (nucleus) clip in preference to the content of the satellite clips. This will tend, as constraints are applied, towards a presentation focussed on a key communicative goal.

For our evaluation, we looked at sets of weightings that covered both of these cases. Within a clip or sequence we chose a relative weighting scheme that favoured background content over elaboration, designed to suit a 'first-time' audience:

$$1 \geq w_i(preparation) \geq w_i(conclusion) > w_i(background) > w_i(elaboration) > w_i(summary) > 0,$$
$$where\ i \in \{clip, segment\}$$

A different precedence may be appropriate for a different audience or presentation conditions – for example, after the initial viewing of a *presentation* the background may be less important and the *summary* may be more important. By choosing a different weighting scheme for audiences with different prior knowledge, it should be possible to tailor an appropriate selection strategy.

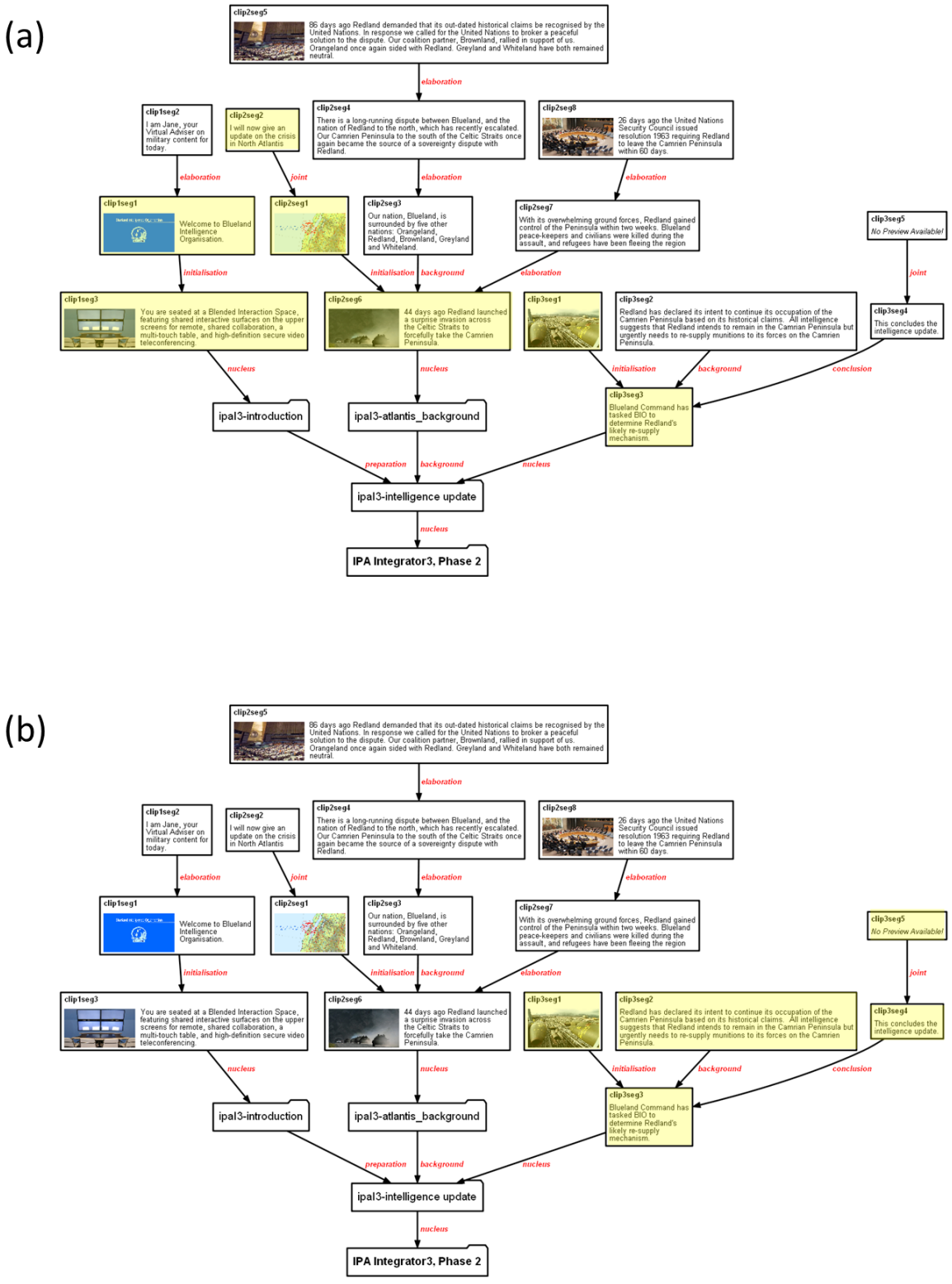


Figure 21: Different weighting schemes will favour different parts of the discourse structure of a presentation. Here, the highlighted elements are favoured for a) $w_{clip}(R) > w_{segment}(R) \ \forall R$, and b) $w_{clip}(R) < w_{segment}(R) \ \forall R$.

To evaluate the effectiveness of this approach, different weighting schemes following these constraints were used to assign weights to the rhetorical relations within the multimedia presentation document, and the scores for each multimedia segment calculated[25]. The set of presentations generated by varying a selection threshold, and discarding content with a score lower than this threshold, was generated and assessed for each of the weighting schemes used. After some experimentation, two weighting schemes were identified as showing promise with the example presentation, and will be discussed further.

7.2 Focussed Selection Strategy

This weighting scheme adopted the approach, illustrated in Figure 21b, where

$$w_{clip}(R) < w_{segment}(R) \;\; \forall R \in \{R\}$$

The values of the weights used were chosen to provide a clear decoupling between the clip and segment structure as the selection threshold is varied.

Table 4: Weighting scheme chosen to favour the clip representing the presentation 'nucleus'.

	Nucleus	Joint	Initialisation	Preparation	Conclusion	Background	Elaboration	Summary
w_{clip}	1	1	1	0.8	0.7	0.6	0.5	0.4
$w_{segment}$	1	1	1	0.98	0.97	0.96	0.95	0.94

In the absence of another quantitative measure, a count of the number of clips and segments contained within the multimedia presentation was obtained as the selection threshold was varied, and normalised against the total number of clips and segments in the original multimedia document, to provide the graph shown in Figure 22. This metric is somewhat indicative of the relative duration of the presentation.

This shows how, as the selection threshold is increased, entire clips would be progressively dropped to meet the presentation constraints, giving 11 possible (hopefully coherent) presentations. The selection thresholds where the generated presentation changes are summarised in Table 5.

[25] A groovy script was written to perform these calculations. This was later incorporated into the IMMP prototype discussed in §8

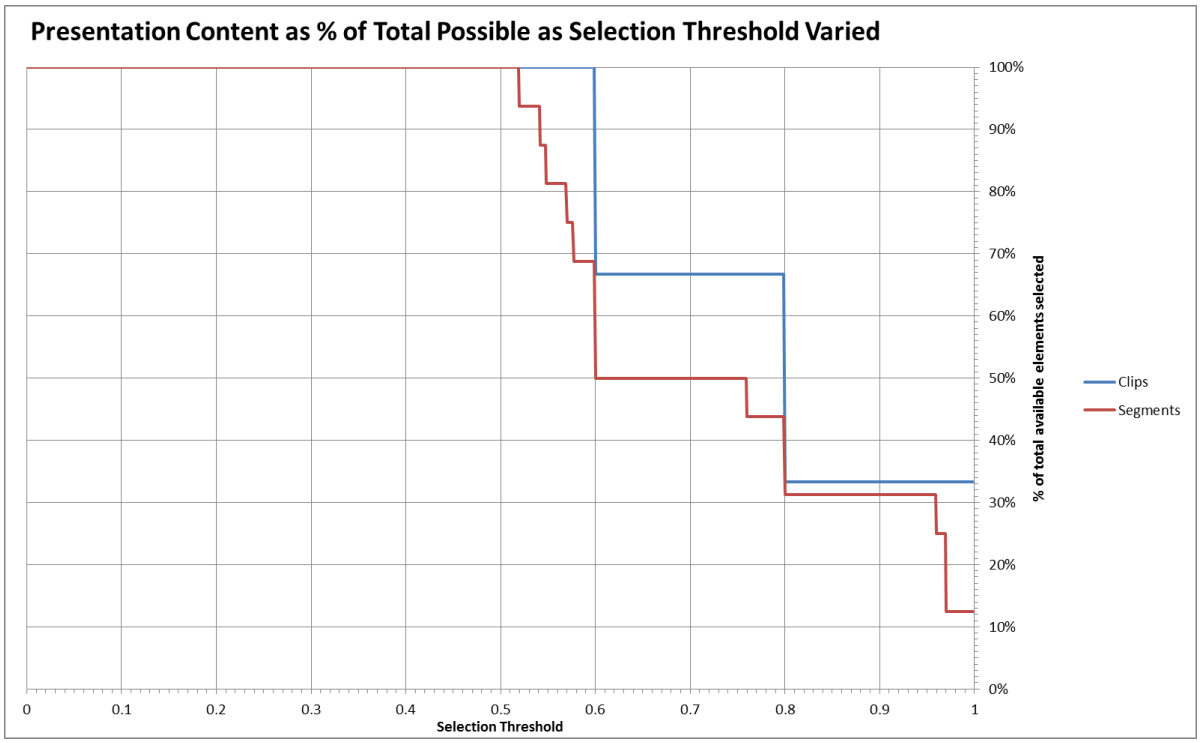


Figure 22: Quantitative evaluation of percentage of multimedia elements contained in multimedia presentation as the selection threshold is varied for the 'focussed' selection strategy.

Table 5: Number of multimedia elements contained in presentation produced at different selection thresholds using the 'focussed' selection strategy.

threshold	0.000	0.520	0.542	0.548	0.570	0.577	0.600	0.760	0.800	0.960	0.970
clips	3	3	3	3	3	3	2	2	1	1	1
segments	16	15	14	13	12	11	8	7	5	4	2

To assess the coherence of the resulting presentations qualitatively, HTML storyboards were generated for each of these 11 possible presentations, and are shown in Appendix D.1. This strategy does indeed generate a set of progressively shorter presentations that are both formally coherent, and subjectively coherent.

7.3 Overview Selection Strategy

This weighting scheme adopted the approach, illustrated in Figure 21a, where

$$w_{clip}(R) > w_{segment}(R) \;\; \forall \, R \in \{R\}$$

The values of the weights used were again chosen to provide a clear decoupling between the clip and segment structure as the selection threshold is varied.

Table 6: Weighting scheme chosen to favour the presentation of each clip's communicative sub-goal.

	Nucleus	Joint	Initialisation	Preparation	Conclusion	Background	Elaboration	Summary
w_{clip}	1	1	1	0.98	0.97	0.96	0.95	0.94
$w_{segment}$	1	1	1	0.8	0.7	0.6	0.5	0.4

Again, a count of the number of clips and segments contained within the multimedia presentation was obtained as the selection threshold was varied, and normalised against the total number of clips and segments in the original multimedia document, to provide the graph shown in Figure 23. This shows how, as the selection threshold is increased, segments for all of the clips are progressively dropped while retaining the overall clip structure, giving 11 possible presentations. The selection thresholds where the generated presentation changes for this weighting scheme are summarised in Table 7.

Table 7: Number of multimedia elements contained in presentation produced at different selection thresholds using the 'overview' selection strategy.

threshold	0.000	0.145	0.240	0.289	0.480	0.490	0.677	0.600	0.700	0.960	0.980
clips	3	3	3	3	3	3	3	3	3	2	1
segments	16	15	14	13	12	11	10	9	7	4	2

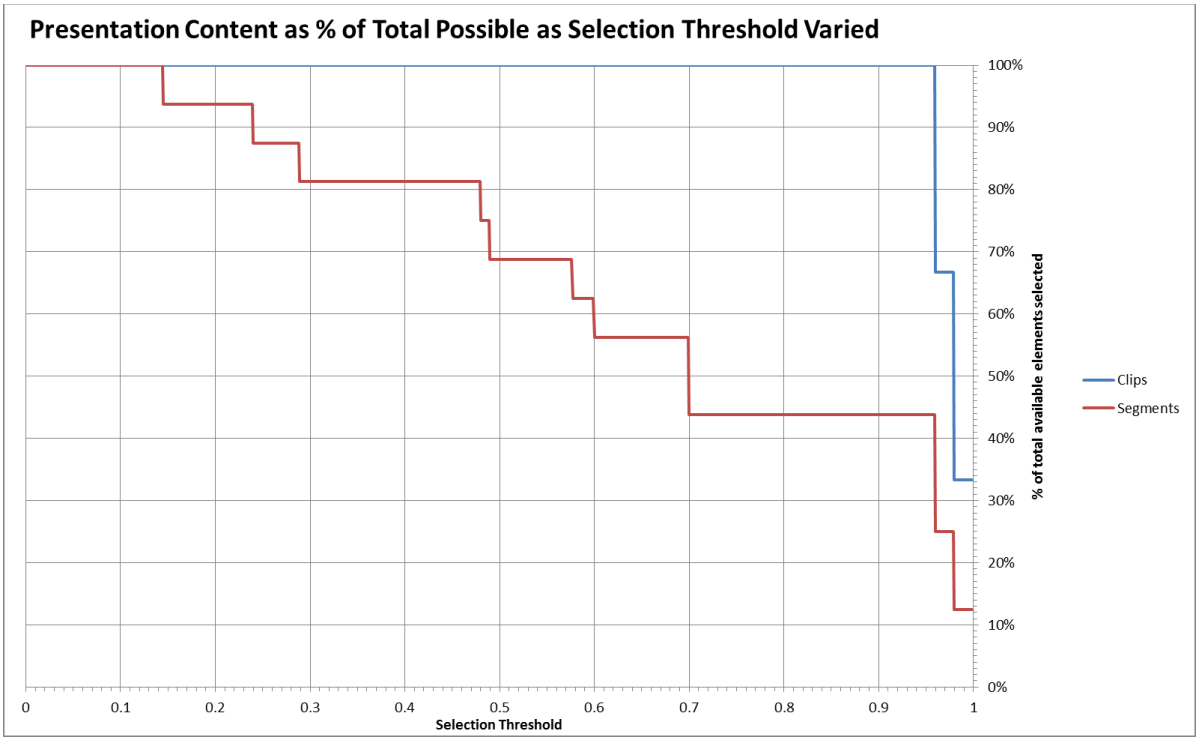


Figure 23: Quantitative evaluation of percentage of multimedia elements contained in multimedia presentation as the selection threshold is varied for the 'overview' selection strategy.

To assess the coherence of the generated presentations qualitatively, HTML storyboards were generated for each of these 11 possible presentations, and are shown in Appendix D.2. Again, this strategy does indeed generate a set of progressively shorter presentations that are both formally coherent, and subjectively coherent.

7.4 Presentation Constraints

The constraints imposed when presenting a multimedia presentation will determine how content needs to be selected from a presentation document. The simplistic weighting approach taken here can be used to address the three constraints discussed earlier:

1. **Topic:** relevant content, and the corresponding rhetorical relations, is determined by the topic to be presented.
2. **Prior Knowledge:** different weighting schemes for the rhetorical relations can be applied when selecting content, based on the prior knowledge of the audience. For example, the 'focussed' weighting scheme discussed in §7.2 favours the main communicative sub-goal of the presentation and it's directly relevant context, and so is appropriate for an audience already cognizant with the situation and the peripheral context. In contrast, the 'overview' weighting scheme discussed in §7.3 favours a narrative structure that includes all of the communicative sub-goals represented within the clips, and so is appropriate for an audience unfamiliar with the situation and requiring more context.
3. **Duration:** the duration of a multimedia presentation will depend on the amount of content to be presented. While there is not a one to one relationship between the number of segments presented, and the time taken to present them, with the approach taken there will be a monotonic relationship with time[26] – as the number of multimedia segments contains within a presentation decreases, the duration of the presentation will also decrease.

In order to explore the relationship between presentation duration and selection threshold more fully, the time taken to present each of the presentation variations using the Virtual Adviser[27] was measured, for both of the weighting schemes considered.

[26] This is not necessarily true in general, but applies in our case because the deeper branches of the discourse tree will be progressively pruned from the presentation.
[27] Using the IMMP prototype discussed in §8

Table 8: Time to present the presentation variations for the 'focussed' selection strategy.

Threshold	Segments	Measured time to give presentation (secs)				Mean (secs)	Stddev (secs)
0.000	16	161.956	161.955	161.909	162.002	161.96	0.04
0.521	15	135.283	135.283	135.267	135.252	135.27	0.01
0.543	14	122.939	122.924	122.877	122.923	122.92	0.03
0.549	13	105.938	105.970	105.970	105.923	105.95	0.02
0.571	12	91.939	91.985	91.892	91.860	91.92	0.05
0.578	11	81.189	81.220	81.219	81.220	81.21	0.02
0.601	8	62.079	62.079	62.032	62.110	62.08	0.03
0.761	7	56.079	56.204	56.110	56.095	56.12	0.06
0.801	5	33.391	33.485	33.422	33.391	33.42	0.04
0.961	4	14.187		14.219	14.187	14.20	0.02
1.000	2	8.734	8.823	8.891	8.875	8.83	0.07

Table 9: Time to present the presentation variations for the 'overview' selection strategy.

Threshold	Segments	Measured time to give presentation (secs)				Mean (secs)	Stddev (secs)
0.01	16	162.034	16.1987	161.956	161.940	161.98	0.04
0.146	15	135.283	135.283	135.251	135.268	135.27	0.02
0.241	14	122.923	122.892	122.939	122.971	122.93	0.03
0.29	13	105.955	105.970	105.970	105.939	105.96	0.01
0.481	12	91.892	91.830	91.891	91.861	91.87	0.03
0.491	11	85.845	85.938	85.830	85.892	85.88	0.05
0.578	10	75.219	75.158	75.188	75.094	75.16	0.05
0.601	9	55.969	60.329	55.922	55.954	57.04	2.19
0.701	7	50.595	50.610	50.564	50.579	50.59	0.02
0.961	4	31.500	31.485	31.484	31.501	31.49	0.01
1	2	8.782	8.985	8.907	8.812	8.87	0.09

These results verify that the relationship between the number of multimedia segments within the presentation (as the selection threshold is varied) and the duration of the presentation is indeed monotonically increasing. Given that the relationship between the selection threshold and the number of multimedia segments monotonically decreases (see Figure 24 and Figure 25), the relationship between the presentation duration and selection threshold is also monotonically decreasing. This means that for any given duration greater than the minimum (determined by the duration of the nucleus segment of the nucleus clip), it is possible to find a selection threshold above which all generated presentations will have a smaller duration. Thus, this weighting scheme can be used to find the longest duration presentation that meets a nominated duration constraint, provided it is greater than or equal to the minimum possible duration of the presentation, and that the duration of a multimedia segment can be estimated reliably.

Both of the weighting schemes explored in this work appear to provide feasible sets of presentations from a source document, as a selection threshold parameter is varied. The weighting scheme chosen determines the relative importance of different structural and narrative elements in the presentation. It is likely that different weighting schemes could be developed to focus on different narrative aspects, to allow the presentation to be further tailored to meet different audience requirements and prior knowledge – for example, placing greater weight on summary relationships and less on background relationships.

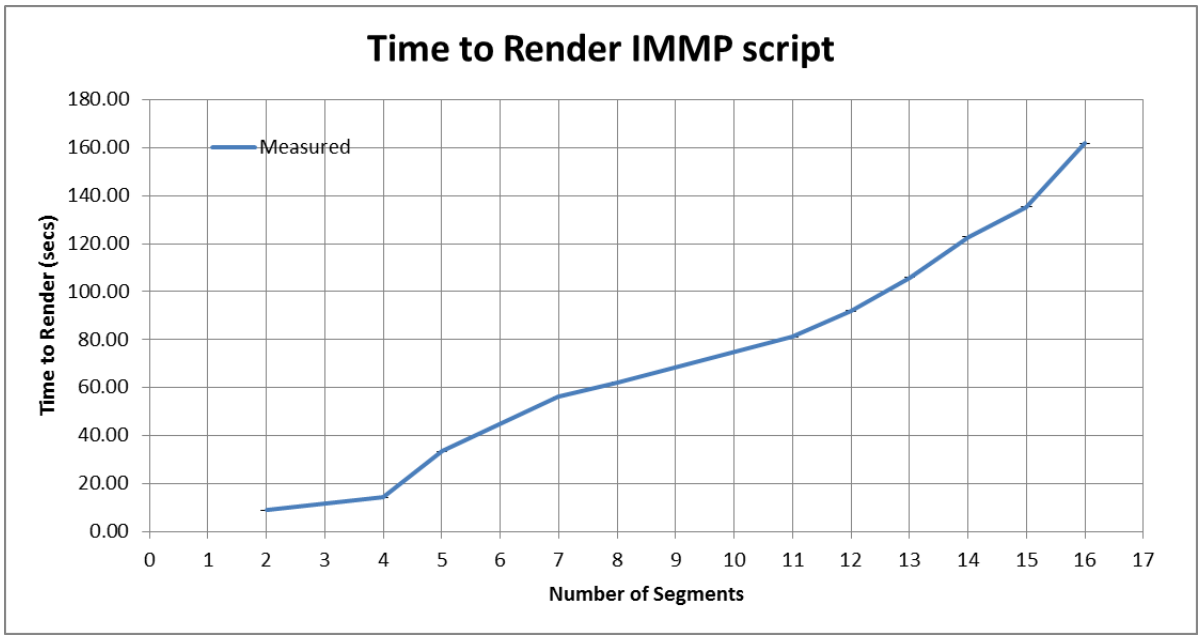


Figure 24: Time to give presentation as the number of multimedia segments within it is varied using the 'focussed' selection strategy.

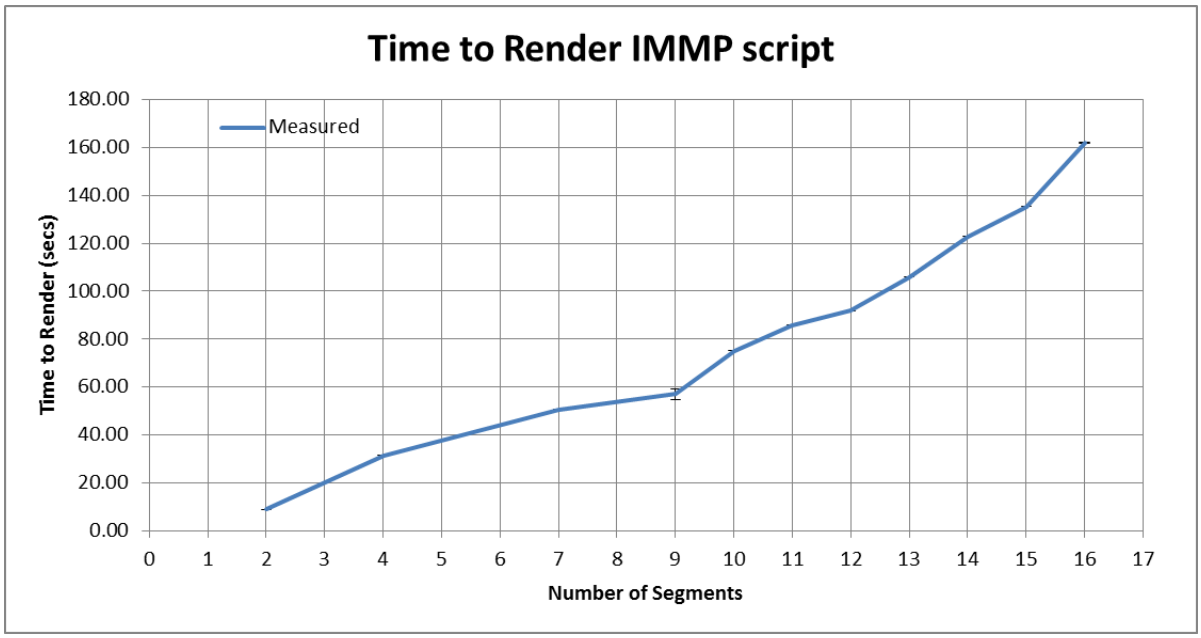


Figure 25: Time to give presentation as the number of multimedia segments within it is varied, using the 'overview' selection strategy.

This approach has been used to develop an IMMP prototype system that takes a multimedia presentation document and selects which content to present given constraints on the presentation duration, which will be discussed further in §8.

8. The IMMP Prototype

Given the demonstrated feasibility of the weighting scheme for rhetorical relations discussed in §7, at producing coherent presentations from a human-authored multimedia document under different presentation constraints, a prototype IMMP system was developed to implement this approach. This system is a long way from achieving the stated goal of an automated capability, but it will allow us to explore the feasibility of this approach with a wide variety of real-life use cases.

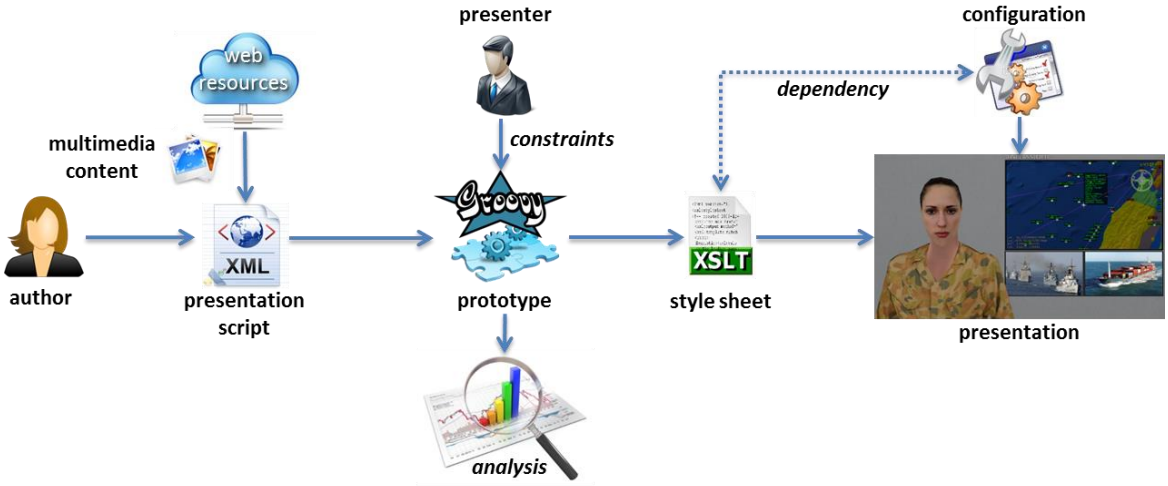


Figure 26: Prototype IMMP system that takes a human-authored multimedia document and presents it given specified constraints.

The workflow for this system is summarised below:

1. A human author generates an XML document[28] describing the multimedia presentation, ensuring:
    a. multimedia content (images, videos, etc.) is stored in accessible web resources such as a wiki
    b. multimedia content for one or more nominated topics is arranged in multimedia segments, clips, and sequences
    c. multimedia content is assigned to appropriate rendering channels and styles, which at this stage are abstract constructs only.
    d. appropriate rhetorical relations for the nominated topics are assigned to the multimedia clips and segments.
2. The XML document is read by a Groovy script which:
    a. parses the XML document

[28] In future this will be done using a web-based authoring tool, but at the moment any XML editor is suitable.

b. assigns weights to the rhetorical relations based on the selected scheme

c. scores each multimedia segment in the document based on this weighting scheme

d. estimates the rendering time for each multimedia segment and clip

e. finds the lowest threshold score (which corresponds to the most included content) that gives a presentation with a duration less than a nominated constraint (if it can be satisfied), and generates a XML presentation using this selection threshold.

f. streams this generated multimedia presentation to the Virtual Adviser system for further processing and rendering.

3. An XSLT script invoked by the Virtual Adviser maps the generated XML presentation in the THML format for execution by the Virtual Adviser. This does three things:

   a. Extracts any implicit content (discussed later)

   b. Processes any timing information (discussed later)

   c. Maps the content in the presentation to (abstract) rendering channels (and potentially styles).

4. The Virtual Adviser processes the THML generated by the XSLT script, and handles the realisation of the generated multimedia presentation using a predefined THML configuration script, that maps how the abstract channels and styles in the generated multimedia presentation are realised, and the layout used in the Virtual Adviser scene. This configuration file constitutes the design of the presentation. Note that the XSLT script and the THML configuration file need to define consistent channels and styles.

5. The Virtual Adviser execution system coordinates the scheduling of the multimedia content, and renders it into the Virtual Adviser scene.

8.1 XML Requirements

In order for the multimedia document to be compatible with the prototype IMMP system, some requirements need to be met:

1. **Topics:** The author may optionally associate one or more topics with the clips and rhetorical relations in the document. If a topic is not specified, a 'default' topic will be allocated to the contents of the document.

2. **Rhetorical Relations:** The author must identify the nucleus of a clip sequence, and the nucleus of each clip. The rhetorical relations of satellite clips and segments must be identified, from the set:

   - **joint**
   - **initialisation**
   - **preparation**

- **conclusion**
- **background**
- **elaboration**
- **summary**

3. **Channels:** The author needs to select an appropriate channel for the multimedia content included in the presentation. The channel should be one of:
    - **narration** – text to be uttered by a Virtual Adviser
    - **caption** – text to be displayed on the screen with the Virtual Adviser
    - **icon** – a graphic or image to be displayed on the screen
    - **monitor** – image or video to be displayed in a virtual 'video monitor'

    While other channels can be used in the document, appropriate changes will need to be made to the configuration of the Virtual Adviser to support these channels.
4. **Multimedia content:** The author must represent multimedia document in one of two ways in the document, using the following tags:
    - **text** – text can be included directly within the document
    - **uri** – other content such as images, videos, or scripts can be referenced by a URI that references a web resource holding the content. File locations can also be included here, but it is not recommended as it is not portable.
5. **Timing:** The author may optionally specify limited timing information for rendering content in the prototype:
    - **delay** – specifies the time delay (in seconds) to be applied before the specified content is rendered. Note that, as the Virtual Adviser handles content scheduling in the prototype, for simplicity the *delay* attribute will only have an effect when used with the *narration* channel. Finer grained timing with other channels can be achieved by breaking content up into separate segments.
    - **duration** – specifies the minimum time (in seconds) to display the specified content. Note that in most cases the time content is displayed also depends on how long it takes for an associated utterance to complete.
6. **Titles:** The author may optionally specify titles with sequences and clips, which are treated as implicit content that is also rendered in the Virtual Adviser scene.
7. **Variables:** The author may optionally define variables (see Appendix B) in the document. It is often useful to define a variable that holds information such as the root location of multimedia content, so that the presentation can be easily updated if this content is moved to another location.

For further information on the XML format, consult Appendix B. The example script shown in Appendix C also provides a useful example of a suitable multimedia presentation.

8.2 IMMP Script

The IMMP prototype was developed as a Groovy script, which evolved for a system originally developed to analyse how well our weighting scheme performed with an example multimedia document. It contains several features aimed at supporting this analysis, but the features relevant to its use as an IMMP system are described in the following sections.

8.2.1 Topics

The script will only process those clips contained in the multimedia document with the specified topic and ontology, and will select the rhetorical relations appropriate for that topic and ontology. If no clips within the document are tagged with this topic and ontology, then no output will be generated. If the topic and ontology is not specified, a default topic and ontology is assumed – only those clips with either a value of 'default' for both the topic and ontology, or none specified, will be processed.

8.2.2 Prior Knowledge

There are two types of presentations that can be generated from the multimedia document, using the different weighting schemes discussed in §7.2 and §7.3 (respectively):

- **Focussed:** the nucleus of the sequence is treated as the most important discourse element, and other content will only be included to support this when time permits. This is deemed to be more appropriate for audiences familiar with the context of the presentation.
- **Overview:** the overall clip structure is treated as more important than the detailed content in each clip, and additional content in each clip will only be included when time permits. This is the default type of presentation generated, as it is considered to better represent the needs for an audience unfamiliar with the context of the presentation.

8.2.3 Maximum Duration

The maximum duration desired for a generated presentation can be specified to constrain the content within the presentation. By default, the script will generate a presentation containing all content relevant to the specified topic and ontology. In this case, the type of presentation based on prior knowledge is not relevant.

In order to simplify the IMMP system, a simple processing pipeline approach was adopted that does not rely on rendering information (such as rendering time) to be fed back to the script. This requires that the script estimates the rendering time for a multimedia segment, based on the optional timing information and a simple heuristic based on the string length of an utterance. This is sufficiently accurate in most cases, and as it was expected that presentation durations would not need to be accurate to less than of order 10 seconds, was deemed to be sufficient for our purposes.

As previously discussed in §7.4, because the relationship between selection threshold and presentation duration is monotonic, it is possible to find either a unique 'longest duration' presentation that meets the specified duration constraint, or no suitable presentation. In the latter case, no presentation is generated.

8.2.4 Output

There are three output options available to the script:

- An XML presentation document can be generated that contains only the multimedia content meeting the specified constraints. In this case, elements are annotated with the weighted score assigned to each multimedia clip and segment.
- A HTML storyboard can be generated representing the content in the generated presentation (see, for example, the storyboards contained in Appendix D).
- A THML stream with embedded XML content can be generated and sent to the specified Virtual Adviser. By default, this tries to connect to a Virtual Adviser running on the local host, but Virtual Adviser services running on other systems can be specified.

8.3 XSLT script

An XSLT script could, in general, convert the XML generated by the IMMP script into any multimedia format, such as SMIL or MHEG, for rendering. In the prototype system it is used to generate THML commands for execution by the Virtual Adviser, based on the structure and content of the generated document. It abstracts content from design by populating THML variables using content found in the document. This allows the realisation of these channels to be specified purely in the Virtual Adviser configuration file. Some of the key functions of the XLST script are:

- It extracts the sequence titles and clip titles to build an implicit 'title' channel for each clip
- It initialises the content of the channels at the start of each clip.
- It populates variables for each channel with content extracted from the multimedia segment. In the prototype system it will only overwrite the content in a channel if there that channel is used (i.e. channel content is retained between segments unless specifically updated).
- It extracts the timing information for each segment and populates corresponding variables.
- It invokes the execution of the multimedia content in a segment, after all content for that segment has been processed, using a macro defined in the Virtual Adviser configuration script.

The XLST script needs to be revised if a modified set of channels have been used in the XML presentation document, so that corresponding variables are initialised. The script can also be revised to produce different execution behaviour.

8.4 Virtual Adviser Layout

The Virtual Adviser THML configuration script defines how the 1 implicit ('title') and 4 explicit channels are realised in the Virtual Adviser scene. It also defines the macros used to initialise each clip, and play to content of each segment using the content defined in the variables populated by the XSLT. If any additional channels are included in the presentation document, how they are realised in the Virtual Adviser scene needs to be defined in this file. Note that any channels used that are not defined both here and in the XSLT script will simply not be rendered.

The default channel layout is shown in Figure 27. It is a relatively simple matter to change how these channels appear in the Virtual Adviser scene by editing the THML configuration script – no other changes are required to the IMMP system in this case.

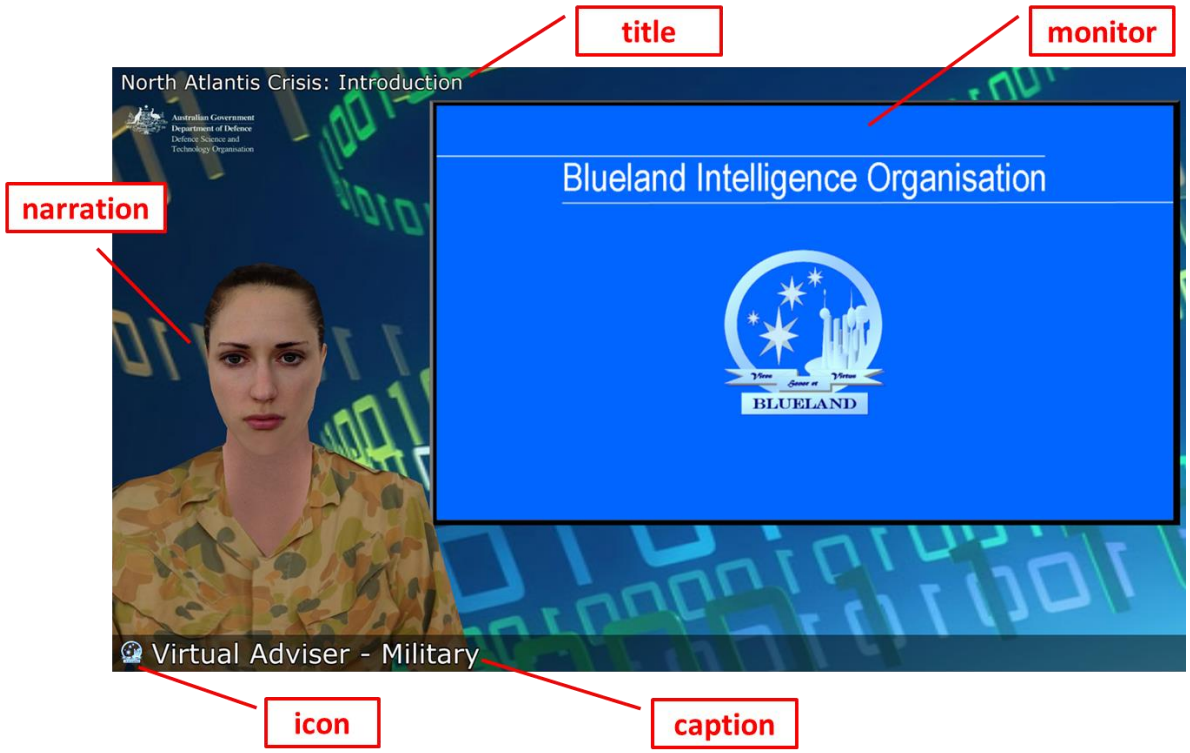


Figure 27: Default layout for multimedia presentations produced by the IMMP prototype, and how content in channels is mapped to the Virtual Adviser scene.

8.5 Packaging

The IMMP prototype system, including the groovy script, XSLT, and THML configuration has been bundled as is available from the JOAD Decision Sciences software repository as the 'source.zip' artefact of the 'dsto.immp.scripts' projects at:

http://c2-maven.dsto.defence.gov.au/nexus

It requires that Java 7 and the Virtual Adviser software be installed. More information on how to get this is available at from the DSTO intranet at:

http://logwiki.dsto.defence.gov.au/display/va

For more information on obtaining, installing, and using the IMMP prototype refer to Appendix E.

9. Conclusion

For this work we have focussed on how we can provide a multimedia narrative capability, exploiting some of the strengths of DSTO's Virtual Adviser capability, and complementing its weaknesses. However, while this has been the focus of the work, it could be easily generalised for other multimedia presentation systems. The approach we have taken is based around human authoring of a multimedia presentation, allowing the domain knowledge and narrative expertise of the author to address complexities such as content generation, selection, and coordination of multimedia elements to achieve the communicative goal. The role of the multimedia presentation capability developed is to select a subset of the content of the multimedia presentation while maintaining coherence to achieve the communicative goal given operational constraints – such as time available or prior knowledge of the audience.

In this system, the multimedia presentation is stored as an XML document, with multimedia segments grouped into re-usable multimedia clips. Each clip can relate to one or more topics. Multimedia clips and segments are related to each other by a simplified set of 7 rhetorical relations, describing the narrative structure of the presentation. By using a simple weighting scheme based on the structural and rhetorical relationships of the content, and the contextual knowledge of the audience, we have shown that, for a realistic example, this approach maintains coherence both formally and subjectively. Building on this we have developed a prototype system for playback of human-authored multimedia presentations using the Virtual Adviser. By default, the content that can be included within the presentation is styled on television news services.

The prototype is intended to provide an initial example of an IMMP system, so that these techniques can be explored with a wider variety of use cases. The prototype system has only demonstrated the feasibility of this approach, based on a single, albeit militarily relevant, example. It still needs to be established how effective this approach is with other examples and scenarios, and work is ongoing to evaluate this system with other scenarios. Similarly, the weighting scheme used was developed based on the single example. It is likely that, with a larger corpus of example, the weights used could be refined. It may be possible in this case to identify sets of weightings that provide more that the two styles of presentation implemented here. Finally, with a greater corpus of examples it is likely that an extended set of rhetorical relations will be required to maintain coherence under different presentation constraints. It will be useful to identify where, as the corpus of examples increases, the simple weighting approach used here breaks down and another approach is needed.

Independently of the avenues of work discussed above, additional work is planned on developing a web-based editor for creation of multimedia presentations that allows rhetorical relations to be assigned to the content included, and saved in the XML format described here. While an initial implementation could simply constrain the rhetorical relations to those suitable for our IMMP prototype, it would also be useful to provide hints or templates for authors unfamiliar with rhetorical relations, and similarly provide hints or templates on how to effectively juxtapose and coordinate multimedia content. We envisage that the editor could also be used to construct re-usable multimedia clips on

nominated topics, for potential re-use by an automated IMMP system. The web-based editor could also potentially provide a corpus of annotated presentation examples from which an automated system could learn how to construct multimedia content and assemble it into a coherent presentation. As an early example of this, we intend to explore how image/graphics and their captions could be stored, retrieved, and modified to suit a particular topic based on a given topic ontology.

The simple weighting scheme discussed in this report may also be of value for automatically generated content (Dall and Donnelly, 2014). By assigning rhetorical relations to the potential discourse elements, and assigning the appropriate weights, different versions of a presentation could be either automatically generated, or selected after generation, to suit the presentation constraints. In this case, different discourse elements could be generated to suit different constraints, so that content might not be repeated in different versions. This adds more complexity to the generation/selection process that will need to be considered.

10. References

Andre, E. (2000) "The Generation of Multimedia Presentations". In: Dale, R., Moist, H. and Somers, H. (eds.) *Handbook of Natural Language Processing*. New York, NY, Marcel Dekker, Inc. 314-338

Andrews, T., Broughton, M. and Estival, D. (2006) "Implementing an Intelligent Multimedia Presentation Planner using an Agent Based Architecture". In: *International Conference on Intelligent User Interfaces,* Sydney, Australia: 29 January

Bal, M. (2009) *Narratology: Introduction to the Theory of Narrative, Third Edition*. Canada, University of Toronto Press

Blanchette, M. (2005) *Military Strikes in Atlantis - A Baseline Scenario for Coalition Situation Awareness*. TR-C3I-TP1-1-2005, [TTCP Technical Report] TTCP C3I Group

Bordegoni, M., Faconti, G., Feiner, S., Maybury, M. T., Rist, T., Ruggieri, S., Trahanias, P. ,Wilson, M. (1997) "A Standard Reference Model for Intelligent Multimedia Presentation Systems". *Computer Standards & Interfaces* **18** 477-496

Buchanan, M. C. and Zellweger, P. T. (2005) "Automatic Temporal Layout Mechanisms Revisited". *ACM Transactions of Multimedia Computing, Communications, and Applications* **1** (1) 60-88

Bui, V., Abbass, H. and Bender, A. (2010) "Evolving Stories: Grammar Evolution for Automatic Plot Generation". In: *2010 IEEE Congress on Evolutionary Computation (CEC),* Barcelona: 18-23 July 2010, IEEE

Colineau, N. and Paris, C. (2003) *Framework for the Design of Intelligent Multimedia Presentation Systems: An architecture proposal for FOCAL*. CMIS Technical Report 03/92, CSIRO

Dall, I. and Donnelly, B. (2014) *Event Sequencing for Situation Narratives*. (In Review),

Ekman, P. and Friesen, W. V. (1977) *Facial Action Coding System*. Pao Alto, U.S.A., Consulting Psychologists Press Inc.

Endsley, M. R. (1995) "Toward a Theory of Situation Awareness in Dynamic Systems". *Human Factors* **37** (1) 32-64

Estival, D., Broughton, M., Zschorn, A. ,Pronger, E. (2003) "Spoken Dialogue for Virtual Advisers in a Semi-Immersive Command and Control Environment". In: *4th SIGdial Workshop on Discourse and Dialogue,* Sapporo, Japan: 5-6 July 2003

Kohler, H., Lambert, D. A., Richter, J., Burgess, G. ,Cawley, T. (2013) "Implementing soft fusion". In: *Information Fusion (FUSION), 2013 16th International Conference on*: 9-12 July 2013

Lambert, D. A. (1999) *Proceedings of the 1999 Workshop on Defense Applications of Signal Processing*, La Salle IL., USA, Lindsey, A., et al. (eds.)

Laney, D. *3D Data Management: Controlling Data Volume, Velocity and Variety*. 6 February 2001. (2001) [Accessed 2013; Available from: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Mann, W. C. and Thompson, S. A. (1988) "Rhetorical Structure Theory: Toward a Functional Theory of Text Organisation". *Text* **8** (3) 243-81

Mayer, R. E. and Moreno, R. (2003) "Nine Ways to Reduce Cognitive Load in Multimedia Learning". *Educational Psychologist* **38** (1) 43-52

McKeown, K. R. (1985) "Discourse Strategies for Generating Natural Language Text". *Artificial Intelligence* **27** 1-41

Nowina-Krowicki, M., Zschorn, A., Pilling, M. ,Wark, S. (2011) "ENGAGE: Automated Gestures for Animated Characters". In: *Australasian Language Technology Association Workshop,* Canberra, ACT, Australia: 1-2 December

Nugent, K. M. (2012) *Development of a Conceptual Design for Intelligent Multimedia Presentation using Model Based Systems Engineering*. University of South Australia, University of South Australia

Paris, C., Colineau, N. and Estival, D. (2004) "Intelligent Multi Media Presentation of information in a semi-immersive Command and Control environment". In: *Australasian Language Technology Workshop,* Sydney

Taboada, M. and Mann, W. C. (2006) "Rhetorical Structure Theory: Looking Back and Moving Ahead". *Discourse Studies* **8** (3) 423-59

Taplin, P., Fox, G., Coleman, M., Wark, S. ,Lambert, D. (2001) "Situation Awareness Using a Virtual Adviser". In: *Talking Head Workshop, OZCHI 2001,* Fremantle, Australia

Wark, S., Lambert, D., Nowina-Krowicki, M., Zschorn, A. ,Pang, D. (2009) "Situational Awareness: Beyond Dots on Maps to Virtually Anywhere". In: *SimTecT 2009,* Adelaide, Australia: 15-19 June 2009

Wark, S. and Lambert, D. A. (2007) "Presenting The Story Behind The Data: Enhancing Situational Awareness Using Multimedia Narrative". In: *MILCOM 2007,* Orlando, FL., IEEE

Wark, S., Zschorn, A., Broughton, M. ,Lambert, D. (2004) "FOCAL: A Collaborative Multimodal Multimedia Display Environment". In: *SimTecT 2004 - Simulation - Better Than Reality?,* Canberra, Australia: 24-27 May

Wark, S., Zschorn, A., Perugini, D., Tate, A., Beautement, P., Bradshaw, J. M. ,Suri, N. (2003) "Dynamic Agent Systems in the CoAX Binni 2002 Experiment". In: *6th International Conference on Information Fusion (Fusion 2003),* Cairns, Australia: 8-11 July 2003

Appendix A: Talking Head Markup Language (THML)

THML is a simple text mark-up language for controlling the Virtual Adviser System. A table of all the available THML commands and their usage is provided below.

A.1. THML Commands

Command	Description
<!-- Comment string -->	Even though untagged text sent to THConsole is ignored, using this tag is the standard way to add comments to marked-up text. A unique and useful feature of these comment tag pairs is that they can enclose existing THML in a script, enabling sections to be 'commented out' during testing etc. (Two things to note: this tag **must** be closed with '-->'; and comments tags can't be nested, as the inner closing dashes-bracket would signify the end of the comment.)
<block>	The previous command is concluded before scheduling the next. For example, `<express happy 0.8 1><block><say>Hi</say>` will wait 1 second (until the end of the `<express>` animation) before speaking the text. (Note: there is an implicit `<block>` between `<say>` statements.)
<brow [both] | [left right [onset] [offset]]>	The character's eyebrows transition to the specified state over *onset* seconds starting at *offset* seconds from the insertion point. If not specified *onset* and *offset* default to 0. A single value will set the position of both brows; two values will set the position of the left brow and right brow independently. The brows are raised and lowered with positive and negative values, respectively. The valid range is [-1, 1]. See `<express>` for a description of valid value formats, and the use of the optional *onset* and *offset* values.
<centre heading pitch roll>	Sets the rotation of the head in degrees that all head motions are relative to.
<default tagName [value1] [value2] ... [valueN]>	Specifies the default values used for the singular event, tagName. (Currently, the only option for tagName is 'wink' but this could be extended in the future.) As values are parsed from the tag they override the default values used by tagName. The default `<wink>` is internally defined as `<default wink 0 1 0 0.2 0 0.2>` (Meaning, left eye open, right eye closed, offset = 0, attack = 0.2, sustain = 0, release = 0.2) (see the definition of `<wink>`). If it was desirable to have the default `<wink>` be with the left eye rather than the right, then the command `<default wink 1 0>`would overwrite only those values (leaving the shape of the wink unchanged). There is no way to unset overwritten values, other than to redefine the original defaults or restart THConsole.
<dequeue [id+]>	This removes the specified *id*(s) from the playback queue. If no *id*(s) are specified then all queued animations are removed.
<echo string>	String is scheduled to be echoed back to the console. This is useful for allowing applications to identify when particular animation sequences have been completed. (Bear in mind that it might be desirable to issue a `<block>` command before this to ensure that the echo occurs at the completion of the scheduled animation (see `<block>`).
<echo! string>	String is echoed back to the console at the time of processing. Contrast this with <echo>, which echoes at the scheduled animation time.)
<engage [on | off | [host [port]]]>	This command can either enable or disable the use of Automatically generated (appropriate) non verbal behaviours. The <enage> command can be used in many forms: `<engage on>`

Command	Description

	enable the use of the ENGAGE system
	`<engage off>`
	
	disables the ENGAGE system
	`<engage localhost 7070>`
	
	enables the ENGAGE system using the specified host and port settings

<engage *command*>	Sends the specified command to the ENGAGE system. Example: `<engage command debug on>`

<express [expression1 value1 | [expression1Value 1] [expression2 value2 | [expression2Value 2] ... [onset] [offset]>	The character transitions to the facial expression, *expression*, with magnitude *value* over time *onset*, starting at *offset* seconds from the insertion point. Multiple attribute/value pairs can be specified to define a blended expression. Valid values of expression are:
	`happiness | happy`
	`sadness | sad`
	`anger | angry`
	`fear | afraid`
	`surprise | surprised`
	`disgust | disgusted`
	`neutral | normal`
	`contempt`
	
	Any unrecognised expression is treated as neutral.
	value **should** be within range [0,1] (but other values, +ve and –ve, provide some interesting effects!)
	Absolute values are the default. Relative values are specified by appending '#'. Percentage changes are specified by appending '%'. For example, if the current value is 0.75, '-0.5#' sets it to 0.25, and then '+50%' sets it to 0.375.
	When *onset* and/or *offset* are not given, the default value of '0' is used. Outside of a `<say>` statement this would result in an immediate effect. (Any other *onset* or *offset* values would be relative to the current time if THConsole was used interactively, or relative to the last `</say>` or `<block>`if scripted.) Within a `<say>` statement an implied *onset* or *offset* is defined by its inter-word position, and any specified offset (positive or negative) is added to this.
	Examples of use:
	`<!-- Example 1: Start expression before 'alive' and -->`
	`<!-- be at value 0.8 in 0.1 seconds -->`
	`<!-- -->`
	`<say>I am<express happy 0.8 0.1>alive</say>`
	
	`<!-- Example 2: Start expression 0.2 seconds after -->`
	`<!-- start to say 'alive' and be at -->`
	`<!-- value 0.8 in 0.1 seconds. -->`
	`<!-- -->`
	`<say>I am<express happy 0.8 0.1 0.2>alive</say>`
	
	`<!-- Example 3: Outside of <say>...</say> (when -->`
	`<!-- there are no word boundaries to -->`
	`<!-- make use of), onset and offset -->`
	`<!-- values define when the changes -->`
	`<!-- occur. In this case, will be happy -->`
	`<!-- in 1 second, afraid in 2 seconds, -->`
	`<!-- and neutral in 3 seconds -->`
	`<!-- -->`
	`<express happy 1 1><express fear 1 1 1><express null 1 1 2>`

<flush>	Aborts any currently spoken utterance and deletes any scheduled animation. The VA is returned to a default pose.
<frown [both] | [left right [onset] [offset]]>	'The character's eyebrows transition to the specified state over *onset* seconds starting at *offset* seconds from the insertion point. A single value will set the position of both brows; two values will set the position of the left brow and right brow independently. The brows are tilted

Command	Description
	downward (indicating deep thought or displeasure) with positive values and tilted upward (indicating worry) with negative ones. The valid range is [-1, 1]. This tag can be used in conjunction with the `<brow>` tag for the desired effect. Note that `<frown>` squeezes the brows together, so that a positive `<frown>` value has a different effect than a negative `<brow>`. See `<express>` for a description of valid value formats, and the use of the optional *onset* and *offset*.
<gesture ...>	*Not yet implemented in VA2*
<groovy url args*>	Executes the specified groovy file or URL and inserts the contents of any println statements back into the input stream. The groovy script with have full read/write access to macros as variables inside the script, hence a macro set with <set name Marcin> will be available as the "name" variable int the groovy script. Arguments passed to the script will be available as the traditional "args" list in groovy. If the script or arguments contain embedded spaces they need to be surrounded by double quotes e.g. "C:\Documents and Settings\blogsj\My Documents\dostuff.groovy". e.g. `<groovy` http://cool/stuff/myfunkyscript.groovy `"This will be the first arg" second_arg arg3 "argument four">`
<help>	One day this tag might display some help messages!
<jaw [<value>] [onset] [offset]>	Opens the jaw to the specified value over with the specified onset and offset. The normal range for *value* is between *0 (closed)* and *1 (fully opened)*.

<load url>	Loads the specified file or URL and inserts its contents into the input stream. This can be a THML file or a plain text file if it is wrapped by suitable tags,
	e.g.
	`<say><load my_speech.txt></say>`
	
	file can be specified as either a file in the current directory, or as relative or absolute paths,
	e.g.
	`<load ../../temp/foo>`
	
	or
	`<load c:/temp/foo>`
	
	The files referenced by `<load>` commands can contain `<load>` statements themselves.
	
	(Beware of recursive/circular references!)

<loadxml xslt_url+ xml_url>	Transforms the specified XML file or URL using specified XSLT transforms and then inserts its contents into the input stream. Parameters for the transforms can be specified as a url query string of the form: xslt_url?param=value&param=value*. URLs or filepaths with embedded spaces mustbe quoted with double quotes e.g. "C:\Documents and Settings\dude\Desktop\some file.xml"
	e.g.
	`<loadxml` http://host/some/path/apple.xslt
	`http://someotherhost/some/other/path/passionfruit.xml>`
	`<loadxml C:/some/path/apple.xslt?colour=red&quantity=10`
	`http://someotherhost/some/other/path/passionfruit.xml>`

<log *mode*>	Changes the logging mode. Where *mode* is one of: `none`, `error`, `warn`, `info`, `debug`, `all` defaults to the value of the property `console.log.level` specified in the `TalkingHead.properties` file or `error` if the property is not specified. This command is not case sensitive and will accept most reasonable synonyms for the *mode* string (e.g. `warning` for `warn`)

Command	Description
<look [yaw [pitch [onset] [offset]]]>	The character's eyes move to the specified *yaw* and *pitch* angles (in degrees) over *onset* seconds starting at *offset* seconds from the insertion point. A single value will set the *yaw* (left-right); two values will set the *yaw* (left-right) and *pitch* (up-down); a third and/or fourth values would specify the *onset* and *offset* respectively. Positive values roll the eyes (to the viewer's) right or up; negative ones roll the eyes (to the viewer's) left or down. The valid ranges are yaw = [-30, 30] and pitch = [-10, 10]. See `<express>` for a description of valid value formats, and the use of the optional *onset* and *offset*.
<maya [offset] command>	Synonymous with `<va ...>`
<maya! command>	Synonymous with `<va! ...>`. This tag was created for the Maya version of the VA. Command can be **any** valid Maya Embedded Language command and can extend over multiple lines, not only allowing the calling of arbitrary pieces of code, but the definition of it as well!
<next>	This plays the next animation in the playback queue.
<pause delay>	Advances the animation insertion point by delay seconds.
<play [id+]>	This plays the specified *id*(s) in the playback queue. If no *id*(s) are specified then it plays all queued animation sequences in order.
<previous>	This plays the previous element in the playback queue.
<queue id [context]*> <commands>+ </queue>	This tag generates the animation specified by the enclosed THML commands but queues it for later playback using the specified *id* rather than play it immediately. This is useful when the animation generation time is an issue and can be pre-generated prior to playback. Optionally a set of context tags can be associated with this queue that can then be used to control which queued animation sequences can and can't be played using the `<select [expression]>` command.
<quit>	Closes the THConsole.
<repeat [id] count>	This repeats either the specified *id* or the current (last) element in the playback queue (if no id is specified) *count* times.
<say>utterance</say>	The character says utterance when scheduled on the animation timeline.
<sayas *wordphonetic_transcription*>	Allows a phonetic transcription of a word to be provided to the TTS for generation. **NOTE:** `sayas` commands are only valid within <say>...</say> statements. Example: `<say>Could you please pass me the <sayas file 'fa&Il></say>`

<script url args*> ... </script>	Provides the input found between the <script>...</script> tags to the script (in groovy this is via the scriptInput variable) and then inserts the contents of any println statements from the execution of the script back into the THML input stream. The groovy script with have full read/write access to macros as variables inside the script, hence a macro set with <set name Marcin> will be available as the "name" variable int the groovy script. Arguments passed to the script will be available as the traditional "args" list in groovy. If the script or arguments contain embedded spaces they need to be surrounded by double quotes
	e.g. "C:\Documents and Settings\blogsj\My Documents\dostuff.groovy".
	
	e.g.
	`<script` http://host/some/path/test.groovy
	`"This will be the first arg" second arg arg3 "argument four">`
	`  <people>`
	`  <person><name>Marcin</name><age>31</age></person>`

Command	Description
	<person><name>Dave</name><age>29</age></person> </people> </script>
<select [boolean_expression]>	Enables queued animations selected by the boolean combination of tags defined by the *boolean expression*. No other queued animations can be played. The boolean expression uses the grammar: <expression> := <term> [|| <term>]* <term> := <factor> [&& <factor>]* <factor> := [!] <tag> | [!] (<expression>) e.g. <queue 1 a b c><say>one</say></queue> <queue 2 a d e><say>two</say></queue> <queue 3 b d e><say>three</say></queue> <select (a || b) && !c> <play 1 2 3> will only say "two, three" Note also that: <select> allows every queued animation to be played <select !> does not allow any queued animation to be played
<send host:port string>	This commands establishes a TCP/IP socket connection to *host* on *port* and sends *string*
<set [macro [definition]]>	Sets macro to represent the specified definition. The definition is then substituted whenever the macro appears within THML (designated by $macro$). This can be useful for easily referencing complex actions. The name for macro can be any number of alphanumeric or underscore characters**, in any order or casing. Macro expansion is also case insensitive, so that any form of casing can be used to make the script more readable. Whitespace that separates macro from definition will be trimmed (as is any on the right of definition), although definition itself can contain whitespace within it. In fact, as well as a macro being defined by another macro, a macro name could be defined by another macro... but why would anyone want to do that?! If definition is not given, then an empty string is stored in macro. This allows a script to continue functioning without errors if the macro is not defined. <set> (with no parameters) will list all currently defined macros and their definitions in alphabetical order. If the characters '<', '>' or '\' are required in the macro definition then they must be escaped with a preceding backslash ('\'). In any place within a THML script that '$' is required to stand as it is (and not be expanded), then it should be escaped. For example... <set fruit bananas> <set like I like $fruit$> ...will set like = 'I like bananas', whereas... <set like I like \$fruit\$> ...will set like = 'I like $fruit', allowing $fruit to be redefined in the future, updating $like at the time it is used. Consider also a possible 'gotcha'... <set 10 ants> <say>I have $10 in my pants</say> ...says 'I have ants in my pants</say>' <say>I have \$10 in my pants</say>' ...says 'I have ten dollars in my pants' Macro detection starts at the leading '$' and extends until the next non-alphanumeric, underscore or

Command	Description

	dollar sign. The trailing '$' is usually optional, but will be necessary when a macro is being butted against other characters of a valid macro name.
	For example,
	`<set f foo>`
	`<echo! $f bar>`
	
	prints 'foo bar'
	`<echo! $fbar>`
	
	prints 'Undefined variable "fbar"'
	`<echo! $f$bar>`
	
	prints 'foobar'
	`<set bar tball>`
	`<echo! $f$bar>`
	
	still prints 'foobar'
	`<echo! $f$$bar> (or <echo $f$$bar$>)`
	
	prints 'football'
	Macro expansion is attempted again at the point of a macro's insertion, for those cases where a macro is defined by other macros; however, beware of the possibility of recursive or circular references!

<sound volume offset url>	This command plays a background sound with a *volume* between 0 (off/min) and 1 (max). The url can be either a local file (absolute or relative path) or a well formed URL such as http://myserver/path/music.wav

<system [offset] command>	Schedule command to be executed by the VA rendering system at *offset* seconds from the insertion point. This tag was created primarily for windowed applications (or scripts that launch them), so all commands are backgrounded by default to allow the VA to freely run. No output is returned to THConsole.
	Multiple commands can be specified in the tag by giving them together on the same line (separated by '&&' for Windows or ';' for Unix), or by entering each command on a new line.
	For example,
	`<system 2.5 calc&&notepad>`
	`<system 2.5 calc notepad>`

<system! command>	As for `<system>` tag, but executes immediately when parsed by THConsole, rather than scheduling on the animation timeline.
	This tag was created for obtaining immediate results from the operating system, and unlike the `<system>` tag it returns its output to THConsole. It is because of this that windowed applications will **block** subsequent markup in the THConsole until the launched application is closed. If the intention was to launch the application without blocking, then command should be backgrounded by prepending with 'start' (for Windows) or appending with '&' (for Unix).
	For example,
	`<system! pwd>`
	`<system! start calc&&notepad>`
	
	launches these applications backgrounded together
	`<system! start calc`
	`start notepad>`
	
	Each line is a new system call, so each line will wait for completion before continuing

<timer *action*>	*action* can be one of *start*, *elapsed* or *reset*: *start* starts the timer and echoes "Timer: started". *elapsed* displays the currently elapsed time in seconds since either *start* or *reset* were called in the following format "Timer: *xxx* seconds" (e.g. "Timer: 4.32 seconds"). *reset* displays the elapsed time since the timer was started with either *start* or *reset* using the same format as *elapsed* and resets the timer back to zero.
<tts command>	Send command to the TTS system immediately when parsed by THConsole. Command needs to

Command	Description
	use the correct syntax for the particular back-end TTS system used (namely, rVoice, or Festival). **rVoice service**: `<tts help>` - displays a list of the available tts commands `<tts voice>` - displays the current tts voice `<tts voice ?>` - displays a list of the available tts voices `<tts voice name>` - sets the tts voice to name, where: name = en_au_f01 sets the Australian female voice = en_rp_m01 sets the British male voice = en sc f01 sets the Scottish female voice = en sc m01 sets the Scottish male voice = en ga f03 sets the US female voice = en ga m01 sets the US male voice = en_ga_m02 sets the alternative US male voice Defaults to en_au_f01. `<tts volume vol>` - sets the volume to parameter vol in the range [0,100] Defaults to 50. `<tts pitch p>` - sets the pitch to parameter p in the range [-10,10] Defaults to 0. `<tts rate r>` - sets the speaking rate to parameter r in the range [-10,10] Defaults to 0. **Festival TTS**: Festival's available commands are rather more limited than those for rVoice. Only `voice` and `scheme` are available. However, the `scheme` command allows arbitrary *Scheme* code to be sent to the Festival TTS server (over multiple lines within the tag). (Scheme is the interpreted LISP dialect that Festival is partly coded and mostly configured with.) This low-level access to Festival comes at a cost: poorly-formed code (such as unclosed parentheses) could upset subsequent speech synthesis. Scheme/LISP code is built upon the concept of 'list manipulation', with lists being demarked by parentheses, and the overall structure consisting of nested lists. Care needs to be taken so that all opening parentheses are balanced by closing ones by the end of the Scheme command, otherwise the command will be left on the Festival server in an unevaluated state. If speech production is failing after sending suspect Scheme code that makes complex use of parentheses, then sending additional right parentheses with `<tts scheme )))))>` might help. Useful Scheme commands for Festival are: `<tts scheme current-voice>` - displays the current voice `<tts scheme (voice.list)>` - displays a list of available voices A voice can be selected by: `<tts voice name>`, where name = cstr_us_awb_arctic_multisyn sets a Scottish male voice cstr_us_jmk_arctic_multisyn sets a Canadian male voice rab diphone sets a low-quality UK male voice don_diphone sets a low-quality UK male voice kal_diphone sets a low-quality US male voice ked_diphone sets a low-quality US male voice
<turn [yaw [pitch [roll [onset] [offset]]]]>	The character's head is turned to the specified *yaw*, *pitch*, and *roll* angles (in degrees) over *onset* seconds starting at *offset* seconds from the insertion point. A single value will set the *yaw* (left-right); a second will set the *pitch* (up-down); a third will set the *roll* (side-to-side); and fourth and/or fifth values will set the *onset* and *offset* respectively. Positive values turn the head to the viewer's right or up; negative ones turn the head to the viewer's left or down. There are no invalid ranges for *yaw*, *pitch* or *roll*. See `<express>` for the use of the optional *onset* and *offset*.
<unset macro>	Removes macro and its definition from the list of macros.

Command	Description

<va [offset] command>	Synonymous with <maya ...> (but created especially for the VA2), this tag will run command on the command port of the rendering system at offset seconds from the insertion point. Whereas the command port of the Maya version of the VA will accept arbitrary definitions or calls to code, the VA2 system only accepts calls to authenticated functions.
	
	Command consists of function calls in the alternative forms of:
	`func1(float1, int1, "string1", etc...);`
	`func2 float2 int2 "string1";`
	
	Semi-colon delimiters are not required for a lone function call, and parameter-less functions can be called with or without empty brackets. String parameters must be surrounded with **double** quotes if they contain embedded whitespace or commas (otherwise they would be parsed as additional parameters).
	
	The current functions are:
	`setAdviser`
	`showAdviser`
	`hideAdviser`
	`viewer`
	`texture`
	`morph`
	
	Function:
	`setAdviser <name>`
	where
	`<name> = jane | gijane | dale | mikeB | gijoe`
	
	Example of usage:
	`<va 2.0 setAdviser jane>`
	
	changes the adviser to 'jane' 2 seconds after the insertion point in the text.
	
	Function:
	`showAdviser`
	`hideAdviser`
	
	Example of usage:
	`<va 2.0 hideAdviser>`
	`<va 5.0 showAdviser>`
	
	hides the Adviser after 2 seconds and then shows it again after 5 seconds from insertion point.
	
	Function:
	`viewer <command>`
	where
	`<command> is a command detailed in the section below`
	
	Example of usage:
	`<va 3.0 viewer window.hide>`
	
	hides the VA2 window 3 seconds after the insertion point.
	`<va viewer window.position=200,500>`
	`<va viewer window.size=800x600>`
	`<va viewer window.title=VA2 rules OK>`
	`<va viewer background.texture=./models/spark.tga>`
	
	sets the VA2 window title, position, size and background.
	
	Function:
	`texture <string>`
	where
	`<string> = 'body=<bodyTexture>' | 'head=<headTexture>'`

Command	Description

	Examples of usage:
	`<va texture body=jacketGrey>`
	changes the body texture to the grey jacket (provided)
	`<va texture head=helen>`
	changes the head texture to the helen texture but does not change the head shape
	
	Function:
	`morph <name> <offset> <value> <duration>`
	
	`where`
	`<name> is the name of the morph target`
	`<offset> is the delay in seconds before ramping up`
	`<value> is the target value (can be absolute e.g. 1.0, relative e.g 0.5# or percentage e.g. 75%)`
	`<duration> the ramp up duration in seconds`
	Examples of usage:
	`<va morph puff 0 1 0.2>`
	blends in the puff morph target to a value of 1 over 0.2 seconds starting immediately

<va! command>	As for the `<va>` tag, but executes immediately when parsed by THConsole, rather than scheduling it on the animation timeline.

<wait start>...<wait until*seconds*>	specifies a timed block of commands, appearing between <wait start> and <wait until *seconds*>, that will be executed the timeline will then block until *seconds* has passed since encountering the <wait start> command.
	e.g. if
	`<timer start>`
	`<wait start>`
	`<say>This is a short sentence.</say>`
	`<timer elapsed>`
	`<wait until 11.25>`
	`<timer elapsed>`
	is executed together.
	The output will be the following:
	`Timer: started`
	`Timer: 3.17 seconds`
	`Timer: 11.27 seconds`
	
	Note: timing resolution can be up to 16 milliseconds, so do not rely on the precision of the timing... e.g. set your atomic clock using these commands

<wink [both] | [left right [offset [attack [sustain [release]]]]]>	The character winks at *offset* seconds from the insertion point. An empty tag inserts a default wink (with the default eye). (See the definition of `<default>`) A single value will wink both eyes; two values allow the eyes to be winked independently (or together). Valid values are 0 or 1 (no wink or wink). Subsequent (optional) values define the *offset* and shape of the wink (in seconds): *offset* defines a delay before the start of the wink; *attack* defines the time to reach the wink value; *sustain* defines the duration the eye is closed; and *release* defines the time taken to open the eye again. Any positive float or integer values are valid.
<xml xslt_url+> ... </xml>	Transforms the XML found between the <xml>...</xml> tags using the specified XSLT transforms and then inserts its contents into the input stream. Parameters for the transforms can be specified as a url query string of the form: xslt_url?param=value&param=value*, e.g. `<xml` http://host/some/path/people2thml.xslt>

Command	Description
	<people> <person><name>Marcin</name><age>31</age></person> <person><name>Dave</name><age>29</age></person> </people> </xml>

A.2. Viewer Commands

Command	Description
<va viewer adviser.hide>	Hides the Adviser
<va viewer adviser.scale=*scale*>	Scales the adviser (A scale of 1.0 represents the default size of the Adviser)
<va viewer adviser.show>	Shows the Adviser
<va viewer axis.hide>	Hides the 3D axis
<va viewer axis.show>	Shows the 3D axis
<va viewer background.hide>	Hides the background image
<va viewer background.move=*x*,*y*,*z*>	Moves the background relatively by *x*, *y*, *z* units where: +*x* is moving to users right +*y* is moving away from the user +z is moving up
<va viewer background.pause>	If the background is a supported video file it will be paused if it is playing
<va viewer background.play>	If the background is a supported video file it will be played
<va viewer background.position=*x*,*y*,*z*>	Positions the background absolutely.
<va viewer background.rewind>	If the background is a supported video file it will be rewound if it is playing
<va viewer background.scale=*width_x_height*>	Scales the current size of the background
<va viewer background.show>	Shows the background image
<va viewer background.size=*width_x_height*>	Sets the size of background
<va viewer background.stop>	If the background is a supported video file it will be stopped if it is playing
<va viewer background.texture=*file*>	Sets the texture image of the background *file* can be a URL or a local file path NOTE: images should use power of two dimensions so that they will work on older systems that don't support NPOTS (non-power of two) textures.
<va viewer background.volume=*volume*>	Sets the desired volume for video files where *volume* is a value between 0 (min) and 1 (max)

Command	Description
<va viewer camera.home>	Resets the Camera to the default home position
<va viewer camera.lookat=*flag*>	Sets the adviser to automatically turn the head towards the camera
<va viewer camera.mouse_and_keyboard=*flag*> <va viewer camera.keyboard_and_mouse=*flag*	Enables or disable camera navigation via mouse and keyboard
<va viewer camera.move=*x*,*y*,*z*>	Sets the camera position relatively. where: +*x* is right, +*y* is into the scene and +*z* is up
<va viewer camera.orientate=*h*,*p*,*r*>	Orientates the camera absolutely. where *h*, *p*, *r* are heading, pitch and roll respectively. Rotations are specified in degrees.
<va viewer camera.position=*x*,*y*,*z*>	Sets the camera position absolutely. where: +*x* is right, +*y* is into the scene and +*z* is up Setting the camera to `0,-30,0` positions the camera infront of the adviser looking directly at the neck
<va viewer camera.reset>	Resets the Camera to the default home position
<va viewer camera.rotate=*h*,*p*,*r*>	Rotates the camera relatively. where *h*, *p*, *r* are heading, pitch and roll respectively. Rotations are specified in degrees.
<va viewer camera.turnto=*flag*>	Sets the adviser to automatically turn body (including head) towards the camera
<va viewer caption.[*id*].backdrop=*flag*>	Sets the backdrop on/off for the on screen caption. The backdrop is useful for providing contrast when the background does not provide enough. (*flag* is true when it is equal to any one of *true*, *on*, *yes*or *1* and false otherwise)
<va viewer caption.[*id*].scroll.speed=*speed*>	Sets scroll speed for on screen caption scrolling where *speed* is treated as a multiplier i.e. a *speed* of `2` makes the text scroll at twice the speed
<va viewer caption.backdrop.colour=*r*,*g*,*b*[,*a*]>	Sets the backdrop colour of the on screen caption
<va viewer caption[.*id*].append=*text*>	Appends to the text of the on screen caption
<va viewer caption[.*id*].backdrop.offset=*offset*>	Allows an offset to be specified for the backdrop. The best value for this depends on the font being used but is usually somewhere between 0.001 and 0.1. Some experimentation is required on the users behalf although the default value is usually reasonable.
<va viewer caption[.*id*].background.colour=*r*,*g*,*b*[,*a*]>	Sets the text background rectangle colour NOTE: that this feature is still in development and is only partially functional
<va viewer caption[.*id*].background.fill=*flag*>	Sets whether or not the text background rectangle fills the screen horizontally (*flag* is true when it is equal to any one of *true*, *on*, *yes* or *1* and false otherwise) NOTE: that this feature is still in development and is only partially

Command	Description
	functional
<va viewer caption[.*id*].background=*flag*>	Sets the text background rectangle on/off (*flag* is true when it is equal to any one of *true*, *on*, *yes* or *1* and false otherwise) NOTE: that this feature is still in development and is only partially functional
<va viewer caption[.*id*].clear>	Removes all items from the caption.
<va viewer caption[.*id*].*command*>	Multiple captions can be used by assigning a numbered *id* (e.g. 1) between the `caption.` and the *command*. For example `<va viewer caption.1.push.text=I'm Caption 1>` and `<va viewer caption.1.hide>`. The only condition on the id is that it is an integer value >= 0. If the specified *id* hasn't been used before a caption for this *id* is created. If the *id* is left off, the default caption is used (e.g. <va viewer caption.push.text=I'm the Default Caption>).
<va viewer caption[.*id*].font=*font*>	Sets the font to use for the on screen caption
<va viewer caption[.*id*].hide>	Hides the on screen caption
<va viewer caption[.*id*].move=*x*,*y*>	Moves the caption relatively to it's current position by *x* pixels across and *y* pixels up where: +*x* is to the right and +*y* is up
<va viewer caption[.*id*].pop>	Pops the last item off of the caption. If the caption is scrolling the item will not be popped until the caption finishes it's current loop. Pop removes gaps automatically so that pop will always remove the first text/image item from the caption.
<va viewer caption[.*id*].position=*x*,*y*>	Positions the caption absolutely *x* pixels across and *y* pixels up where: origin (0,0) is the bottom left corner
<va viewer caption[.*id*].push.gap=*pixels*>	Inserts a gap of the specified *pixel* width to the end of the caption
<va viewer caption[.*id*].push.image=*url*>	Inserts the image specified by the *url* to the end of the caption. If there is already an item on the caption and there is no gap after it, a default gap of 15 pixels will automatically be inserted. If this is not desired the user should call <va viewer caption.push.gap=0> before inserting this item.
<va viewer caption[.*id*].push.text=*text*>	Inserts the specified *text* to the end of the caption. If there is already an item on the caption and there is no gap after it, a default gap of 15 pixels will automatically be inserted. If this is not desired the user should call <va viewer caption.push.gap=0> before inserting this item.
<va viewer caption[.*id*].scroll=*flag*>	Sets scrolling on/off for on screen caption (*flag* is true when it is equal to any one of *true*, *on*, *yes* or *1* and false otherwise)
<va viewer caption[.*id*].show>	Shows the on screen caption
<va viewer caption[.*id*].size=*size*>	Sets the font size of the on screen caption in pixels
<va viewer caption[.*id*].text=*text*>	Clears and then sets the text of the on screen caption. **DEPRECATED** use `caption.push.text` instead
<va viewer caption[.*id*]colour=*r*,*g*,*b*,*a*>	Sets the colour of the on screen caption where: *r*, *g*, *b* and *a* are the red, green, blue and alpha components respectively

Command	Description
	each component is a floating point value between 0.0 and 1.0 inclusive
<va viewer character.lookat.camera=*flag*>	Synonymous with <va viewer camera.lookat=...> Sets the adviser to automatically turn the head towards the camera
<va viewer character.turnto.camera=*flag*>	Synonymous with <va viewer camera.turnto=...> Sets the adviser to automatically turn body (including head) towards the camera
<va viewer cursor.hide>	Hides the mouse cursor
<va viewer cursor.show>	Shows the mouse cursor
<va viewer log.level=*level*>	Sets the **OSG** logging level to the specified level. Where level can be: always, fatal, warn, notice, info, debug, debug_fp
<va viewer monitor.hide>	Hides the monitor image
<va viewer monitor.move=*x*,*y*,*z*>	Moves the monitor relatively by *x*, *y*, *z* units where: +*x* is moving to users right +*y* is moving away from the user +*z* is moving up
<va viewer monitor.pause>	If the monitor is a supported video file it will be paused if it is playing
<va viewer monitor.play>	If the monitor is a supported video file it will be played
<va viewer monitor.position=*x*,*y*,*z*>	Positions the monitor absolutely
<va viewer monitor.rewind>	If the monitor is a supported video file it will be rewound if it is playing
<va viewer monitor.scale=*width_x_height*>	Scales the current size of the monitor
<va viewer monitor.show>	Shows the monitor image
<va viewer monitor.size=*width_x_height*>	Sets the size of the monitor
<va viewer monitor.stop>	If the monitor is a supported video file it will be stopped if it is playing
<va viewer monitor.texture=*file*>	Sets the texture image of the monitor *file* can be a URL or a local file path NOTE: images should use power of two dimensions so that they will work on older systems that don't support NPOTS (non-power of two) textures.
<va viewer monitor.volume=*volume*>	Sets the desired volume for video files where *volume* is a value between 0 (min) and 1 (max)
<va viewer quit>	Shutsdown the viewer
<va viewer texture.load=*file*>	Preloads the specified texture file (this may pause the rendering thread while the texture is loaded) file can be a URL or a local file
<va viewer window.decorated=*flag*>	Sets whether or not the Window should have a frame with a title bar or not (*flag* is true when it is equal to any one of *true*, *on*, *yes* or *1* and false

Command	Description
	otherwise)
<va viewer window.deiconify>	Restores an iconified (minimised) window **Deprecated:** use window.show instead
<va viewer window.hide>	Hides the viewer window completely (i.e. does not minimise to task bar)
<va viewer window.iconify>	Iconifies (minimises) the window to the task/system bar
<va viewer window.ontop=*flag*>	Sets whether or not the Window stays on top of other windows (*flag* is true when it is equal to any one of *true*, *on*, *yes* or *1* and false otherwise) (**NOTE:** this functionality is only partially working under X11)
<va viewer window.position=*x,y*>	Positions the viewer window (in pixel units) origin (0,0) = screens top left corner) +x runs horizontally right across the screen +y runs vertically down the screen
<va viewer window.show>	Show the viewer window. Force the viewer window to the top, so it can be used to place the viewer window above all others if it's already visible. (**NOTE:** currently on X11 the viewer is not forced to the top if it is already visible. This will be fixed in the near future)
<va viewer window.size=*width_x_height*>	Sets the viewer's window size in pixels
<va viewer window.title=*title*>	Sets the window title (Default = *.: VA2 :.*)

Appendix B: XML Specification of Multimedia Document for IMMP

This appendix provides a specification of the XML format for an IMMP presentation and explains how a document should be logically structured.

B.1. IMMP Document Structure

The core structure of the XML immp script is discussed below.

B.1.1 <immp>

This is the primary node chosen for the IMMP document. Originally <script> was used, but this caused confusion with computer language coding systems in some text editors.

The <*immp*> tag may contain:

- an '*id*' attribute
- an optional '*author*' attribute
    - human readable name (example 'Steve Wark'). A comma separated list of names if there are multiple authors (example 'Steve Wark, Marcin Nowina-Krowicki')
- an optional '*email*' attribute
    - a comma separated list of email addresses if there are multiple authors (they should appear in the same order as the 'author' attribute, if used, so that they can be correctly associated)
- an optional '*created*' attribute. This should either be a simple time string using time date format (for example 15:38 15-NOV-2013) or in ISO-8601 format so that it can be easily parsed by a machine.

If generated or pre-processed, it may also contain optional attributes indicative of the processing done:

- '*source*' is the name of the source IMMP document
- '*schema*' is the name of the selection schema applied
- '*threshold*' is the value of the selection threshold applied
- '*topic*' is the name of the topic filter applied
- '*ontology*' is the ontology of the topic filter applied.

Variables (see Variables) can be defined at this level.

The <*immp*> element may contain an optional <*title*> element that can be used as part of the presentation preparation.

B.1.2 <sequence>

An IMMP document must contain one or more sequences, which may or may not relate to similar topics. The IMMP script in this case provides a way of grouping multiple sequences together. Sequences are intended to contain a stand-alone IMMP presentation – you can think of them as like a slide pack.

The <*sequence*> tag should contain:

- An '*id*' attribute (to uniquely distinguish it from other sequences within the script). This could also, depending on the layout, be used as a title or header banner for the sequence.
- An optional '*layout*' attribute that specifies the default layout to apply to the clips within this sequence. If not specified the default for the style document will apply.

The <*sequence*> element can optionally contain a <*title*> element than may be used as part of the presentation preparation.

B.1.3 <clip>

An IMMP sequence must contain one or more clips, which should contain a narratively coherent set of multimedia content suitable for explaining a particular concept. How much content is contained within a clip depends on the stylistic preferences of the (human) author, but you can think of them as like a single slide in a slide pack. In order to support re-use, a clip should be able to stand alone without requiring content from a previous clip.

The <*clip*> tag should contain:

- An '*id*' attribute to uniquely identify it within a sequence (note that for storage and retrieval in a database this may be concatenated with other identifiers to create a globally unique id).
- An optional '*layout*' attribute that specifies the channel layout to apply to this clip. If not specified the default layout for the sequence will apply.

The <*clip*> element should also contain:

- A <*title*> element that provides an implicit preparation element of the clip. How this is rendered will depend on the layout used.
- Optional (zero or more) <*topic*> elements (see Topics) used to support clip storage, retrieval, and re-use.
- One or more <*rst*> elements to support intelligent content selection. The '*topic*' and '*ontology*' attributes for these elements are optional.

B.1.4

An IMMP clip should contain one or more segments, which contain temporally concurrent elements of multimedia content. Segments should be presented in the sequence in which they appear in the clip. The detailed timings for these segments are determined by the layout used for the clip.

The <*segment*> tag should contain:

- An '*id*' attribute to uniquely identify it within the clip (only).

The <*segment*> element should contain:

- An <*rst*> element used for content selection. Note that '*topic*' and '*ontology*' attributes are not required for this tag at this level.

B.1.5 <content>

An IMMP segment contains one or more multimedia content elements which are intended to be presented concurrently.

The <*content*> tag should contain:

- A '*channel*' attribute specifying how (logically) this content is to be rendered
- An optional '*style*' attribute specifying what (logical) formatting, timing and transition effects to apply to this content
- An optional '*duration*' attribute specifying the minimum duration over which to display this content. This is useful for scheduling, for example, how long to display images if there are no other timing constraints within the segment.
- An optional '*delay*' attribute specifying when this content should be displayed from the start of the segment. Note that how (or whether) this attribute is handled depends on the rendering system used. For example, with the Virtual Adviser this attribute is only handled for content intended as narrative.

The <*content*> element should also contain formatting tags that are to be interpreted by the rendering system. Some common examples of tags that may be used are:

```
<!-- for text content -->
<text>this is the text to be displayed</text>

<!-- for online content (e.g. images, video, scripts, etc) -->
<uri>http://thisIsTheURIToGetTheContentFrom</uri>
```

B.1.6 Variables

Variables can be defined within the body of the document to support, among other things, simplification of referencing frequently used content, and specification of the location of content accessed from a file-system or web service. The syntax chosen for this example is similar to that used for XSLT:

```
<!-- defining the variable -->
<variable name="theVariableName">theVariableValue</variable>

<!-- referencing the variable within a tag (e.g. <uri>) -->
<uri>{$theVariableName}</uri>
```

Implicit variables, substituting for content of the XML document, that can be used in the presentation are:

{$script.id}
{$script.title}
{$script.author}
{$script.created}
{$script.source}
{$script.schema}
{$script.topic}
{$script.ontology}
{$script.threshold}

{$sequence.id}
{$sequence.title}
{$sequence.layout}

{$clip.id}
{$clip.title}
{$clip.layout}
{$clip.rst.name}
{$clip.rst.nucleus}

{$segment.id}
{$segment.rst.name}
{$segment.rst.nucleus}

{$content.channel}
{$content.style}
{$content.duration}
{$content.delay}

B.1.7 Topics

In order to retrieve clips (and other content) from a database so that they can be re-used within a dynamically generated presentation, metadata about the topic of the clip needs to be attached to the clip. A particular clip may be apropos to multiple topics, so multiple topic tags are required. The semantics of the topic tags used will depend on an ontology, and hence needs to be an attribute of the topic tag. In this example, the syntax chosen is:

```
<!-- specifying a topic for a clip -->
<topic name="theTopic" ontology="theTopicOntology"/>
```

B.2. Rhetorical Relations

Rhetorical relations apply at different levels of abstraction within a sequence or clip. They could be unary with respect to the respective structural construct (relate to its nucleus), or relate to other components within the (same) construct. Rhetorical relations are optional additions to these constructs.

Used in this way they support winnowing of the presentation to provide appropriate generation/playback of content in response to user requests, dialogue, or preferences. Elements related by the 'joint' relation are considered to be part of the same discourse element for purposes of content selection and processing. Support for rhetorical relations requires a unique ID for these structural elements, and each element may have multiple rhetorical relationships to other elements.

Some elements in a presentation implicitly represent a rhetorical relation. Eg. The <*title/*> elements map to a component of the 'preparation'.

The rhetorical relations used in the XML document are:

Nucleus – the core concept or narrative component for the parent element.

Background – provides background to help understand the nucleus. May not be required if reviewing a presentation after the first play-through.

Summary – this element summarises the nucleus. May be sufficient for an executive summary.

Preparation – sets up or sign-posts narrative for presentation of nucleus, often helps establishes the context.

Initialisation – a special case of *preparation* used at the start of a clip, which should always be included with the clip.

Conclusion – similar to the *preparation*, but sign-posts the completion of the current narrative thread that can support context-switching.

Joint – allows logical grouping of elements within an RST construct

Elaboration – of the nucleus of the parent element, or the nominated sibling

The rhetorical relations are represented by '<*rst*>' tags that relate to topics and ontologies, that are specified within the tag for a clip element. Since segments are always associated with a clip, the *topic* and *ontology* are not required for <*rst*> tags within these elements. Note that where segments may need to take on different rhetorical roles for different topics, the clip should be sub-divided to support this via the <*rst*> tag attributes associated with clips.

The scope of the rhetorical relation is within the containing element only. If the *nucleus* attribute is not specified for an <*rst*> tag, the relation applies to the nucleus of its parent element. Some examples are shown below:

```
<!--  define rhetorical relations for clips in a sequence -->
<!--  defines nucleus clip of the parent sequence -->
```

```
<rst name="nucleus" topic="theTopic" ontology="theOntology"/>

<!--  rhetorical elements that by default relate to the sequence nucleus -->
<rst name="preparation" topic="theTopic" ontology="theOntology"/>
<rst name="initialisation" topic="theTopic" ontology="theOntology"/>
<rst name="conclusion" topic="theTopic" ontology="theOntology"/>
<rst name="background" topic="theTopic" ontology="theOntology"/>
<rst name="summary" topic="theTopic" ontology="theOntology"/>
<rst name="elaboration" topic="theTopic" ontology="theOntology"/>
<rst name="joint" topic="theTopic" ontology="theOntology"/>

<!-- rhetorical elements applying to nominated clips -->
<rst name="elaboration" nucleus="theTargetClipID" topic="theTopic"
    ontology="theOntology"/>
<rst name="joint" nucleus="theTargetClipID" topic="theTopic"
    ontology="theOntology"/>

<!--  defines rhetorical relations for segments of a clip -->
<!--  defines nucleus of the parent clip - topic & ontology not used -->
<rst name="nucleus"/>

<!--  rhetorical elements that by default relate to the clip nucleus -->
<rst name="preparation"/>
<rst name="initialisation"/>
<rst name="conclusion"/>
<rst name="background"/>
<rst name="summary"/>
<rst name="elaboration"/>
<rst name="joint"/>

<!-- rhetorical elements applying to nominated segment -->
<rst name="elaboration" nucleus="theTargetSegmentID"/>
<rst name="joint" nucleus="theTargetSegmentID"/>
```

After pre-processing, one or more <*score*> elements may be added to the <*rst*> tags to support pipeline processing, which includes:

- A '*schema*' attribute specifying the scoring scheme used for processing
- A '*value*' attribute specifying the cumulative score obtained for the discourse element

For example, after pre-processing, an <rst> tag may look like:

```
<!-- rst element after scoring by an IMMP component -->
<rst name="elaboration" nucleus="clip1seg1">
    <score schema="rev1" value="0.855000000000000"/>
    <score schema="rev2" value="0.855000000000000"/>
    <score schema="rev3" value="0.760000000000000"/>
    <score schema="rev4" value="0.490000000000000"/>
</rst>
```

B.3. Stylesheet Document Structure

While not used in the examples in this report, there is a proposed stylesheet document format that specifies how channels are rendered, how they are arranged within a layout, and how the presentation is scheduled. Default channels, layouts, and styles need to be specified within a stylesheet document. The stylesheet is intended to be used with an IMMP service or executive to manage coordination of multimedia channels.
The core structure of the XML stylesheet is discussed below.

B.3.1 <stylesheet>

This is the primary node chosen for the IMMP stylesheet document. The <*stylesheet*> element contains one or more <*layout*> elements defining how channels are to be rendered. Variables can be defined at this level (see Variables)

B.3.2 <layout>

An IMMP stylesheet must contain one or more layouts. The <*layout*> defines how channels are rendered, and the styles that are applied to that channel.

The <*layout*> tag should contain:

- An '*id*' attribute that is referenced by the IMMP script
- An optional '*pause*' attribute that specifies how long to pause (in seconds) between segments.
- An optional '*default*' attribute. At least one of the layout elements defined within the stylesheet must contain a *default* attribute set to '*true*'. This default layout is used whenever the IMMP script references a layout that is not defined within the stylesheet.

B.3.3 <channel>

The layout element must contain one or more <*channel*> elements. Channels specify how content is to be rendered.

The <*channel*> tag should contain:

- An '*id*' attribute that is referenced by the content element within IMMP script.
- A '*viewer*' attribute that provides a symbolic reference to the viewer to be used to render the channel. Valid enumerations of this attribute depend on the rendering system.
- A '*xslt*' attribute that specifies the XLST transform (and any arguments) to apply to the <*content*> element of the IMMP script, to generate the commands to send to the specified viewer application.
- An optional '*init*' attribute that specifies the uri of an initialisation script the viewer application is to apply when the channel is first initialised within the presentation.
- An optional '*cleanup*' attribute that specifies the uri of a clean-up script the viewer application is to apply when presentation completes.
- An optional '*precedence*' attribute that specifies the order in which channels are to be rendered – this may be required if multiple channels apply to the same viewer application. The default precedence is '0' – higher precedence is rendered after lower precedence.

- An optional '*waitForCompletion*' attribute that specifies if the presentation executive should wait for this channel to complete rendering of the content sent to it before rendering the next segment. The default if not specified is '*false*'. If the content in the IMMP script has a specified duration attribute, the presentation executive should wait this long before deeming the segment completed whether or not the *waitForCompletion* attribute is '*true*'.
- An optional '*default*' attribute. At least one of the channel elements defined within the layout must contain a *default* attribute set to '*true*'. This default channel is used whenever the IMMP script references a channel that is not defined within the layout.

B.3.4 <style>

The <*style*> element must contain one or more style elements. Styles specify different ways of presenting content within a channel, and could include effects such as transitions between segments (e.g. fade-in and fade-out) or text styles to use.

- '*Leadin*' style effects are sent to the viewer application prior to rendering of the content passed to the channel.
- '*Leadout*' style effects are sent to the viewer application prior to rendering of the content passed to the channel.

The style tag should contain:

- An '*id*' attribute that is reference by the IMMP content element within the IMMP script.
- An optional '*leadin*' attribute that specifies the uri of a script the specified viewer application is to apply before the content is rendered.
- An optional '*leadout*' attribute that specifies the uri of a script the specified viewer application is to apply after the content is rendered.
- An optional '*delay*' attribute that specifies how long to delay (in seconds) rendering this content.
- An optional '*default*' attribute. At least one of the style elements defined within the channel must contain a *default* attribute set to '*true*'. This default style is used whenever the IMMP script references a style that is not defined within the specified channel.

B.3.5 Variables

As with the IMMP document, variables can be defined within the body of the style document to support, among other things, simplification of referencing frequently used formats or timings. The syntax chosen for this example is similar to that used for XSLT:

```
<!-- defining the variable -->
<variable name="theVariableName">theVariableValue</variable>

<!-- referencing the variable within a channel definition -->
```

```
    <channel id="channelID" renderer="${theVariableName}"/>
```

Appendix C: Example IMMP Document

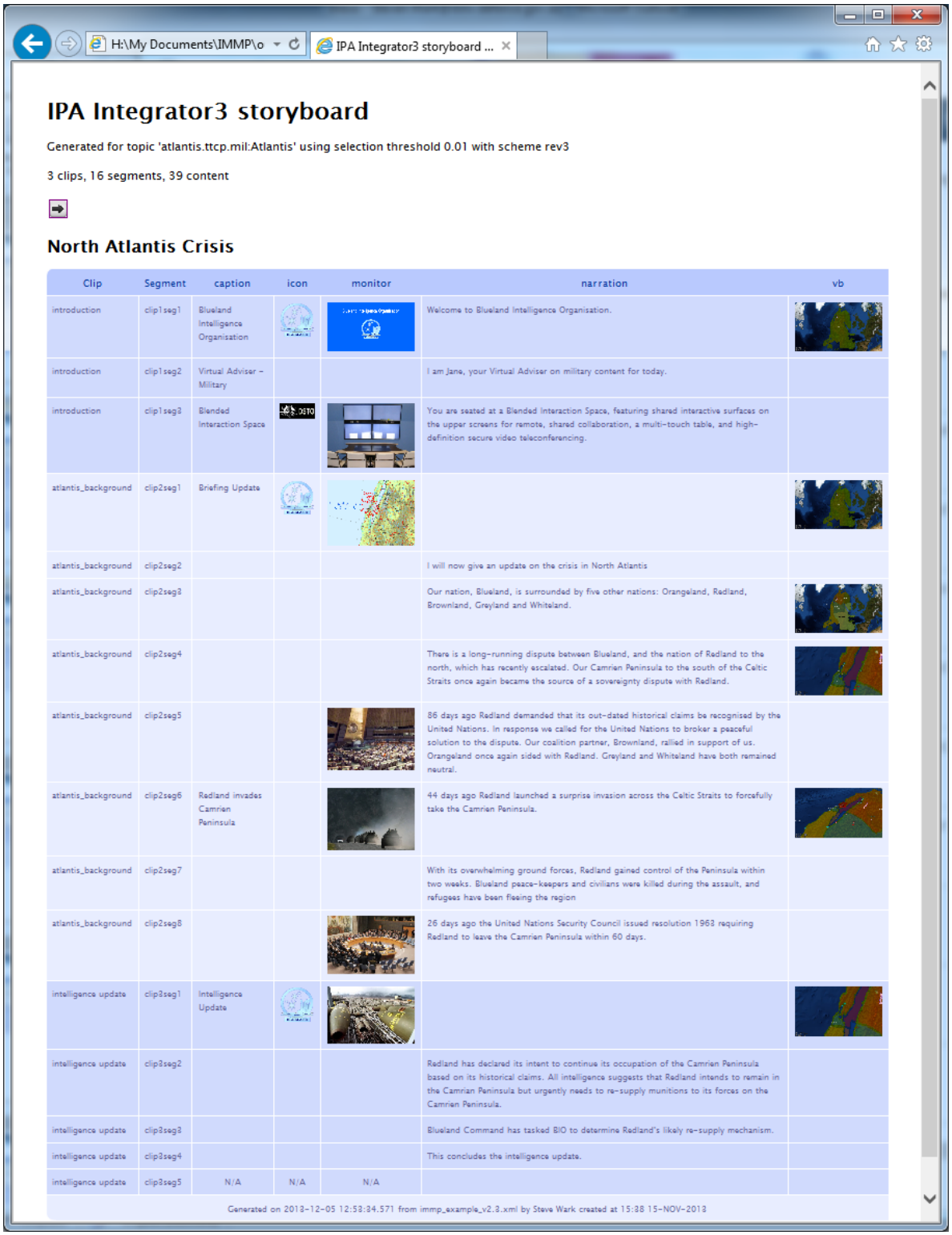
```
<?xml version="1.0" encoding="UTF-8"?>

<!-- ************************************************************************* -->
<!-- Example script based on IPA Integrator3 presentation - modified slightly to    -->
<!-- remove apparent inconsistencies.                                               -->
<!--                                                                                -->
<!-- RST can apply a different levels of abstraction within a sequence, clip, or    -->
<!-- segment. They could be unary with respect to the respective structural const-  -->
<!-- ruct, or in relation to other components within the (same) construct.          -->
<!-- Used in this way they support winnowing of the presentation to provide approp- -->
<!-- iate generation/playback of content in response to user requests, dialogue, or -->
<!-- preferences. An RST element may be composed from multiple structural elements. -->
<!-- Support for RST elements requires unique ID for these structural elements, and -->
<!-- each element may have multiple RST relationships to other elements.            -->
<!--                                                                                -->
<!-- Unary RST relations (wrt structural elements) are:                             -->
<!--    initialisation                                                              -->
<!--    preparation                                                                 -->
<!--    background                                                                  -->
<!--    summary                                                                     -->
<!--    conclusion                                                                  -->
<!--    nucleus                                                                     -->
<!--                                                                                -->
<!-- Within a segment, all rst relations for content can be considered to be unary  -->
<!-- wrt the segment and not other content elements.                                -->
<!--                                                                                -->
<!-- Some elements in a presentation implicitly represent an RST relation. Eg. The  -->
<!-- <title/> elements map to a component of the 'preparation'.                     -->
<!--                                                                                -->
<!-- 24-OCT-2013 Steve Wark, DSTO.                                                  -->
<!-- Initial Version                                                                -->
<!-- ************************************************************************* -->
<immp id="IPA Integrator3" author="Steve Wark" created="15:38 15-NOV-2013">
    <!-- need some way of referencing variables for, e.g, media locations -->
    <!-- this could, in principle, also represent a query to a database to retrieve
         the nominated content -->
     <variable name="media">http://logwiki.dsto.defence.gov.au/download/attachments/60227927</variable>
     <variable name="scripts">http://logwiki.dsto.defence.gov.au/download/attachments/60227927</variable>

     <sequence id="Integrator3, Phase 2" layout="my_favorite">
        <title>North Atlantis Crisis</title>
        <clip id="introduction" layout="this_one_is_better">
            <title>Introduction</title>

            <!-- topic tags for clip -->
            <topic name="Atlantis" ontology="atlantis.ttcp.mil"/>
            <topic name="BIO"      ontology="atlantis.ttcp.mil"/>
            <topic name="BIS"      ontology="dsto.defence.gov.au"/>

            <!-- this clip represents the nucleus for these topics -->
            <rst name="nucleus" topic="BIS" ontology="dsto.defence.gov.au"/>
            <rst name="nucleus" topic="BIO" ontology="atlantis.ttcp.mil"/>

            <!-- this clip represents the preparation for these topics -->
            <rst name="preparation" topic="Atlantis" ontology="atlantis.ttcp.mil"/>

            <!-- implicit segment sequencing assumed here -->
            &lt;segment id="clip1seg1">
                <rst name="initialisation"/>
                <!-- specify time to allow for this content to complete -->
                <content channel="vb" style="reset" duration="3"><uri>{$scripts}/integrator-continent2.vbx</uri></content>
                <content channel="monitor" style="fade_in"><uri>{$media}/bio-background.png</uri></content>
                <content channel="narration"><text>Welcome to Blueland Intelligence Organisation.</text></content>
                <content channel="caption" style="clear"><text>Blueland Intelligence Organisation</text></content>
                <content channel="icon" style="clear"><uri>{$media}/BL8-icon.png</uri></content>
            &lt;/segment>
            &lt;segment id="clip1seg2">
                <rst name="elaboration" nucleus="clip1seg1"/>
                <content channel="narration" style="smile"><text>I am Jane, your Virtual Adviser on military content for
                    today.</text></content>
                <content channel="caption" style="clear"><text>Virtual Adviser - Military</text></content>
            &lt;/segment>
            &lt;segment id="clip1seg3">
                <rst name="nucleus"/>
                <content channel="monitor" style="left_swipe"><uri>{$media}/BIS-overview.jpg</uri></content>
                <content channel="narration"><text>You are seated at a Blended Interaction Space, featuring shared
                    interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and
                    high-definition secure video teleconferencing.</text></content>
                <content channel="caption" style="clear"><text>Blended Interaction Space</text></content>
                <content channel="icon" style="clear"><uri>{$media}/DSTO-cresttext.jpg</uri></content>
            &lt;/segment>
        </clip>

        <clip id="atlantis_background">
            <title>Situation Background</title>
            <topic name="Atlantis" ontology="atlantis.ttcp.mil"/>

            <!-- this clip represents the background for the sequence -->
            <rst name="background" topic="Atlantis" ontology="atlantis.ttcp.mil"/>
```

```
        <!-- for clip re-use should ensure all channels are re-initialised  -->
        &lt;segment id="clip2seg1"&gt;
            <!-- this segment represents part of the preparation for the clip -->
            <rst name="initialisation"/>

            <!-- specify time to allow for this content to complete -->
            <content channel="vb" style="reset" duration="3"><uri>{$scripts}/integrator-continent2.vbx</uri></content>
            <content channel="monitor" style="swipe_left"><uri>{$media}/camrien2-integrator.png</uri></content>
            <content channel="caption" style="clear"><text>Briefing Update</text></content>
            <content channel="icon" style="clear"><uri>{$media}/BL8-icon.png</uri></content>
        &lt;/segment&gt;
        &lt;segment id="clip2seg2"&gt;
            <!-- this segment represents part of the preparation for the clip -->
            <!-- it should be concatenated with the previous preparation segment -->
            <rst name="joint" nucleus="clip2seg1"/>
            <content channel="narration"><text>I will now give an update on the crisis in North
                Atlantis</text></content>
        &lt;/segment&gt;
        &lt;segment id="clip2seg3"&gt;
            <!-- this segment represents part of the background for the clip -->
            <rst name="background"/>
            <content channel="vb" duration="1"><uri>{$scripts}/integrator-borders2.vbx</uri></content>
            <content channel="narration"><text>Our nation, Blueland, is surrounded by five other nations: Orangeland,
                Redland, Brownland, Greyland and Whiteland.</text></content>
        &lt;/segment&gt;
        &lt;segment id="clip2seg4"&gt;
            <!-- this segment represents elaboration for the background segment -->
            <!-- it should be concatenated with the previous background segment -->
            <rst name="elaboration" nucleus="clip2seg3"/>
            <content channel="vb" duration="1"><uri>{$scripts}/integrator-camrien.vbx</uri></content>
            <content channel="narration"><text>There is a long-running dispute between Blueland, and the nation of
                Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic
                Straits once again became the source of a sovereignty dispute with Redland.</text></content>
        &lt;/segment&gt;
        &lt;segment id="clip2seg5"&gt;
            <!-- this segment represents further elaboration of the background for the clip -->
            <rst name="elaboration" nucleus="clip2seg4"/>
            <content channel="monitor"><uri>{$media}/UNGA.jpg</uri></content>
            <content channel="narration"><text>86 days ago Redland demanded that its out-dated historical claims be
                recognised by the United Nations. In response we called for the United Nations to broker a peaceful
                solution to the dispute. Our coalition partner, Brownland, rallied in support of us. Orangeland once
                again sided with Redland. Greyland and Whiteland have both remained neutral.</text></content>
        &lt;/segment&gt;
        &lt;segment id="clip2seg6"&gt;
            <!-- this segment represents the nucleus for the clip -->
            <rst name="nucleus"/>
            <content channel="vb" duration="1"><uri>{$scripts}/integrator-invasion.vbx</uri></content>
            <content channel="caption" style="clear"><text>Redland invades Camrien Peninsula</text></content>
            <content channel="monitor"><uri>{$media}/invasion.jpg</uri></content>
            <content channel="narration"><text>44 days ago Redland launched a surprise invasion across the Celtic
                Straits to forcefully take the Camrien Peninsula.</text></content>
        &lt;/segment&gt;
        <!-- this is an unusual case - rely on rendering engine to insert appropriate pauses between segments -->
        &lt;segment id="clip2seg7"&gt;
            <!-- this segment represents elaboration of the nucleus for the clip -->
            <!-- by default assume that elaboration (etc) refers to the nucleus -->
            <rst name="elaboration"/>
            <content channel="narration"><text>With its overwhelming ground forces, Redland gained control of the
                Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and
                refugees have been fleeing the region</text></content>
        &lt;/segment&gt;
        &lt;segment id="clip2seg8"&gt;
            <!-- this segment represents further elaboration for the clip -->
            <!-- needs to explicitly reference previous elaboration, otherwise would be included as part of
                 elaboration of nucleus -->
            <rst name="elaboration" nucleus="clip2seg7"/>
            <content channel="monitor"><uri>{$media}/UNSC.jpg</uri></content>
            <content channel="narration"><text>26 days ago the United Nations Security Council issued resolution 1963
                requiring Redland to leave the Camrien Peninsula within 60 days.</text></content>
        &lt;/segment&gt;
    </clip>

    <clip id="intelligence update">
        <title>Intelligence Update</title>
        <topic name="Atlantis" ontology="atlantis.ttcp.mil"/>
        <topic name="BIO"      ontology="atlantis.ttcp.mil"/>

        <!-- this clip is the nucleus for this topic -->
        <rst name="nucleus"         topic="Atlantis" ontology="atlantis.ttcp.mil"/>

        <!-- it is the conclusion for this topic -->
        <rst name="conclusion" topic="BIO" ontology="atlantis.ttcp.mil"/>

        &lt;segment id="clip3seg1"&gt;
            <!-- this segment is part of the preparation for the clip -->
            <rst name="initialisation"/>
            <content channel="vb" duration="1"><uri>{$scripts}/integrator-camrien.vbx</uri></content>
            <content channel="monitor" style="swipe_left"><uri>{$media}/munitions.jpg</uri></content>
            <content channel="caption" style="clear"><text>Intelligence Update</text></content>
            <content channel="icon" style="clear"><uri>{$media}/BL8-icon.png</uri></content>
        &lt;/segment&gt;
        &lt;segment id="clip3seg2"&gt;
            <!-- this segment forms the background for the clip -->
            <rst name="background"/>
            <!-- I have broken up the original text from the preceding narrative so as to provide some lead-in as an
                 independent clip -->
```

```
                <content channel="narration"><text>Redland has declared its intent to continue its occupation of the
                    Camrien Peninsula based on its historical claims.  All intelligence suggests that Redland intends to
                    remain in the Camrien Peninsula but urgently needs to re-supply munitions to its forces on the Camrien
                    Peninsula.</text></content>
            &lt;/segment&gt;
            &lt;segment id="clip3seg3"&gt;
                <!-- this segment represents the nucleus for the clip -->
                <rst name="nucleus"/>
                    <content channel="narration"><text>Blueland Command has tasked BIO to determine Redland's likely re-
                        supply mechanism.</text></content>
            &lt;/segment&gt;
            &lt;segment id="clip3seg4"&gt;
                <!-- this segment represents part of the conclusion for the clip -->
                <rst name="conclusion"/>
                <content channel="narration"><text>This concludes the intelligence update.</text></content>
            &lt;/segment&gt;
            &lt;segment id="clip3seg5"&gt;
                <!-- this segment represents part of the conclusion for the clip -->
                <rst name="joint" nucleus="clip3seg4"/>
                <content channel="caption" style="clear"/>
                <content channel="icon" style="clear"/>
                <content channel="monitor" style="clear"/>
            &lt;/segment&gt;
        </clip>
    </sequence>
</immp>
```

Appendix D: Example IMMP Storyboards

D.1. Focussed Selection Strategy



H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

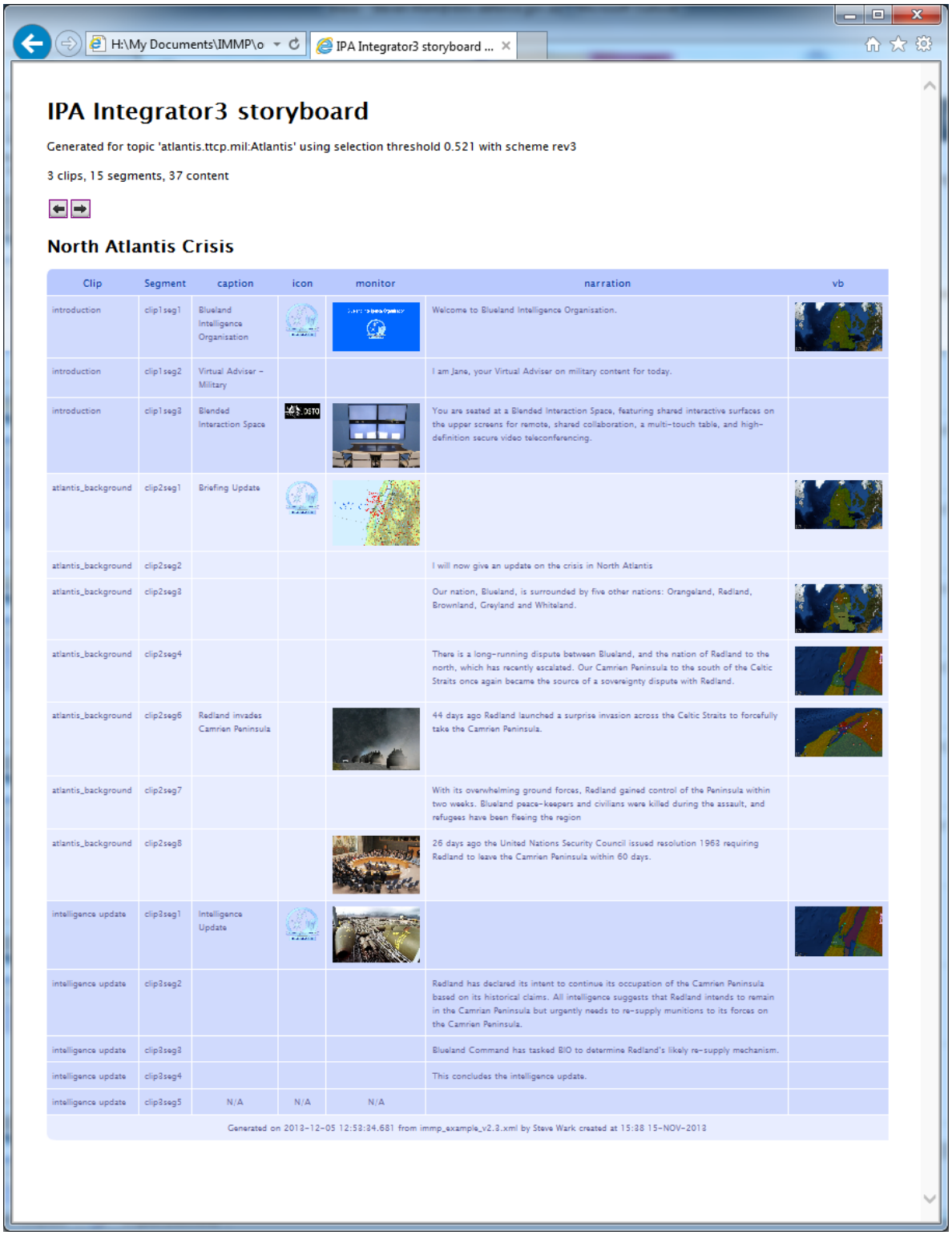
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.01 with scheme rev3

3 clips, 16 segments, 39 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg4				There is a long-running dispute between Blueland, and the nation of Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic Straits once again became the source of a sovereignty dispute with Redland.	
atlantis_background	clip2seg5				86 days ago Redland demanded that its out-dated historical claims be recognised by the United Nations. In response we called for the United Nations to broker a peaceful solution to the dispute. Our coalition partner, Brownland, rallied in support of us. Orangeland once again sided with Redland. Greyland and Whiteland have both remained neutral.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
atlantis_background	clip2seg8				26 days ago the United Nations Security Council issued resolution 1963 requiring Redland to leave the Camrien Peninsula within 60 days.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:34.571 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013



H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.521 with scheme rev3

3 clips, 15 segments, 37 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg4				There is a long-running dispute between Blueland, and the nation of Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic Straits once again became the source of a sovereignty dispute with Redland.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
atlantis_background	clip2seg8				26 days ago the United Nations Security Council issued resolution 1963 requiring Redland to leave the Camrien Peninsula within 60 days.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrien Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:34.681 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.543 with scheme rev3

3 clips, 14 segments, 35 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg4				There is a long-running dispute between Blueland, and the nation of Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic Straits once again became the source of a sovereignty dispute with Redland.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:34.743 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.549 with scheme rev3

3 clips, 13 segments, 33 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser - Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrian Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:34.806 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.571 with scheme rev3

3 clips, 12 segments, 32 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrian Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:34.868 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.578 with scheme rev3

3 clips, 11 segments, 30 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrian Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:34.981 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.601 with scheme rev3

2 clips, 8 segments, 21 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:35.04 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.761 with scheme rev3

2 clips, 7 segments, 19 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrian Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has taskad BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:35.087 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

IPA Integrator3 storyboard

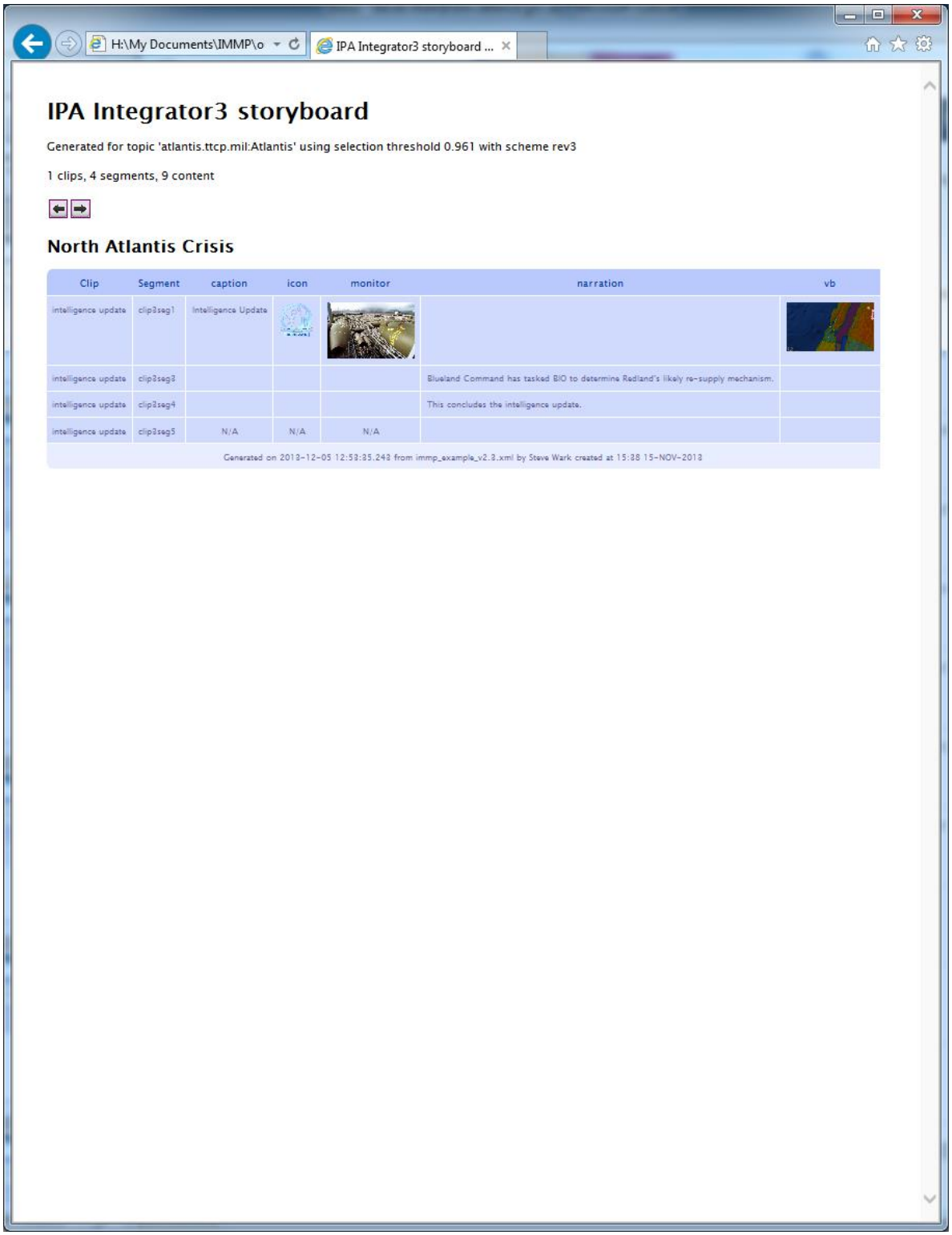
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.801 with scheme rev3

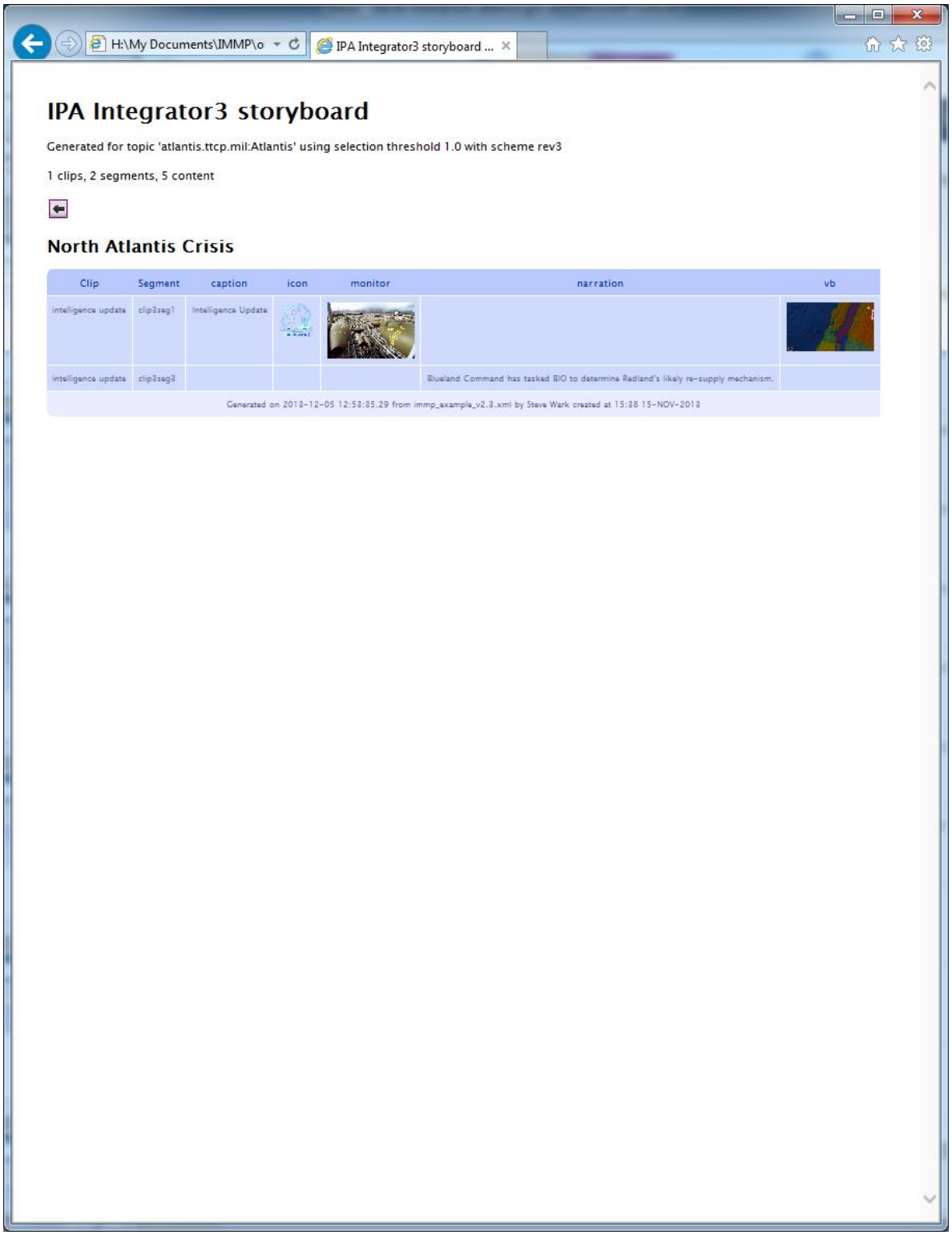
1 clips, 5 segments, 10 content

North Atlantis Crisis

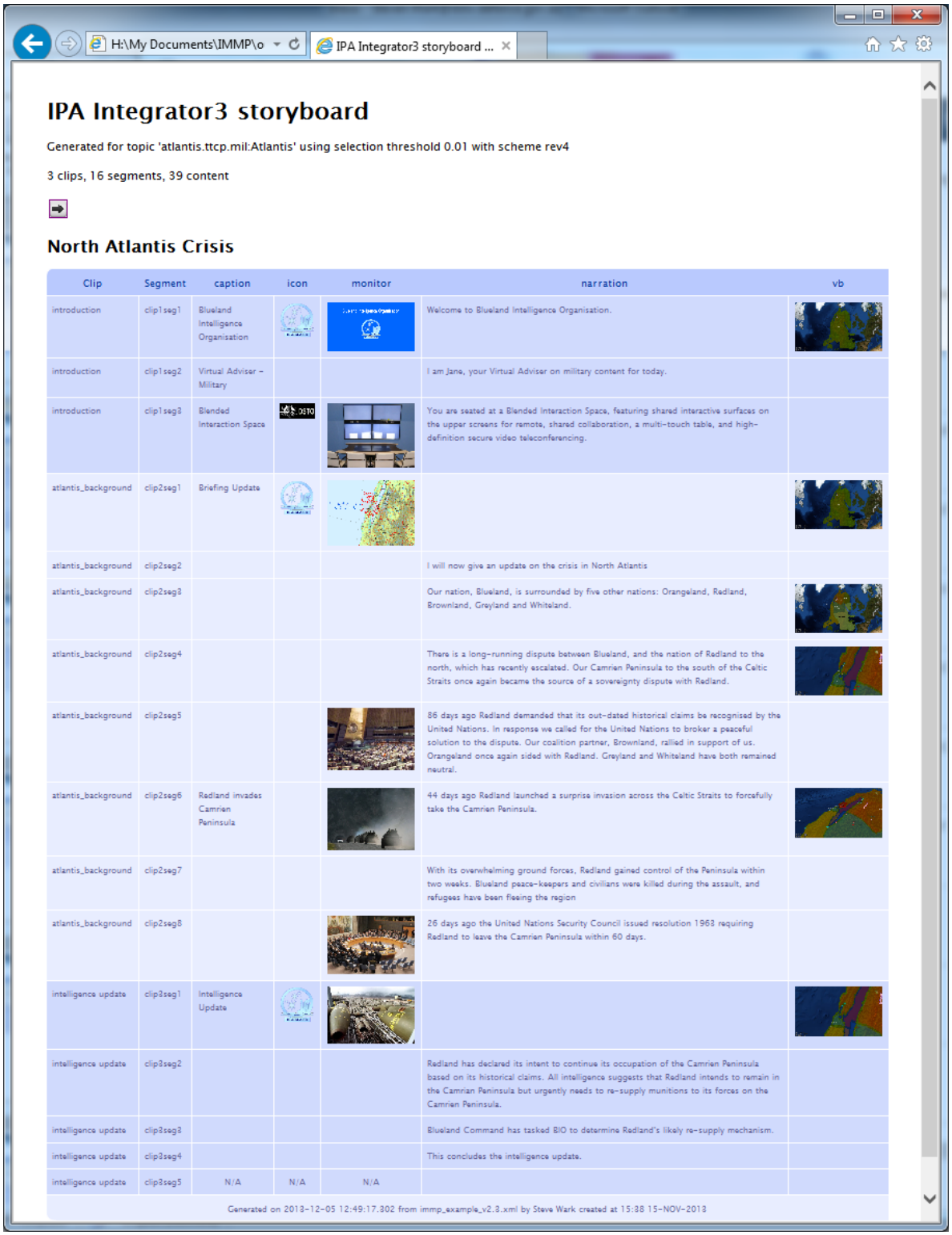
Clip	Segment	caption	icon	monitor	narration	vb
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:53:35.165 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013


H:\My Documents\IMMP\o
IPA Integrator3 storyboard ...
IPA Integrator3 storyboard
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.961 with scheme rev3
1 clips, 4 segments, 9 content
North Atlantis Crisis
Clip
Segment
caption
icon
monitor
narration
vb
intelligence update
clip3seg1
Intelligence Update
intelligence update
clip3seg3
Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.
intelligence update
clip3seg4
This concludes the intelligence update.
intelligence update
clip3seg5
N/A
N/A
N/A
Generated on 2013-12-05 12:53:35.243 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013


H:\My Documents\IMMP\o
IPA Integrator3 storyboard ...
IPA Integrator3 storyboard
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 1.0 with scheme rev3
1 clips, 2 segments, 5 content
North Atlantis Crisis
Clip
Segment
caption
icon
monitor
narration
vb
intelligence update
clip3seg1
Intelligence Update
intelligence update
clip3seg3
Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.
Generated on 2013-12-05 12:53:35.29 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

D.2. Overview Selection Strategy



H:\My Documents\IMMP\o ... IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.01 with scheme rev4

3 clips, 16 segments, 39 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser - Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg4				There is a long-running dispute between Blueland, and the nation of Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic Straits once again became the source of a sovereignty dispute with Redland.	
atlantis_background	clip2seg5				86 days ago Redland demanded that its out-dated historical claims be recognised by the United Nations. In response we called for the United Nations to broker a peaceful solution to the dispute. Our coalition partner, Brownland, rallied in support of us. Orangeland once again sided with Redland. Greyland and Whiteland have both remained neutral.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
atlantis_background	clip2seg8				26 days ago the United Nations Security Council issued resolution 1968 requiring Redland to leave the Camrien Peninsula within 60 days.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:17.302 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.146 with scheme rev4

3 clips, 15 segments, 37 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg4				There is a long-running dispute between Blueland, and the nation of Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic Straits once again became the source of a sovereignty dispute with Redland.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
atlantis_background	clip2seg8				26 days ago the United Nations Security Council issued resolution 1963 requiring Redland to leave the Camrien Peninsula within 60 days.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrien Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:17.584 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.241 with scheme rev4

3 clips, 14 segments, 35 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg4				There is a long-running dispute between Blueland, and the nation of Redland to the north, which has recently escalated. Our Camrien Peninsula to the south of the Celtic Straits once again became the source of a sovereignty dispute with Redland.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:17.724 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.29 with scheme rev4

3 clips, 13 segments, 33 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser - Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
atlantis_background	clip2seg7				With its overwhelming ground forces, Redland gained control of the Peninsula within two weeks. Blueland peace-keepers and civilians were killed during the assault, and refugees have been fleeing the region	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrian Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:17.865 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

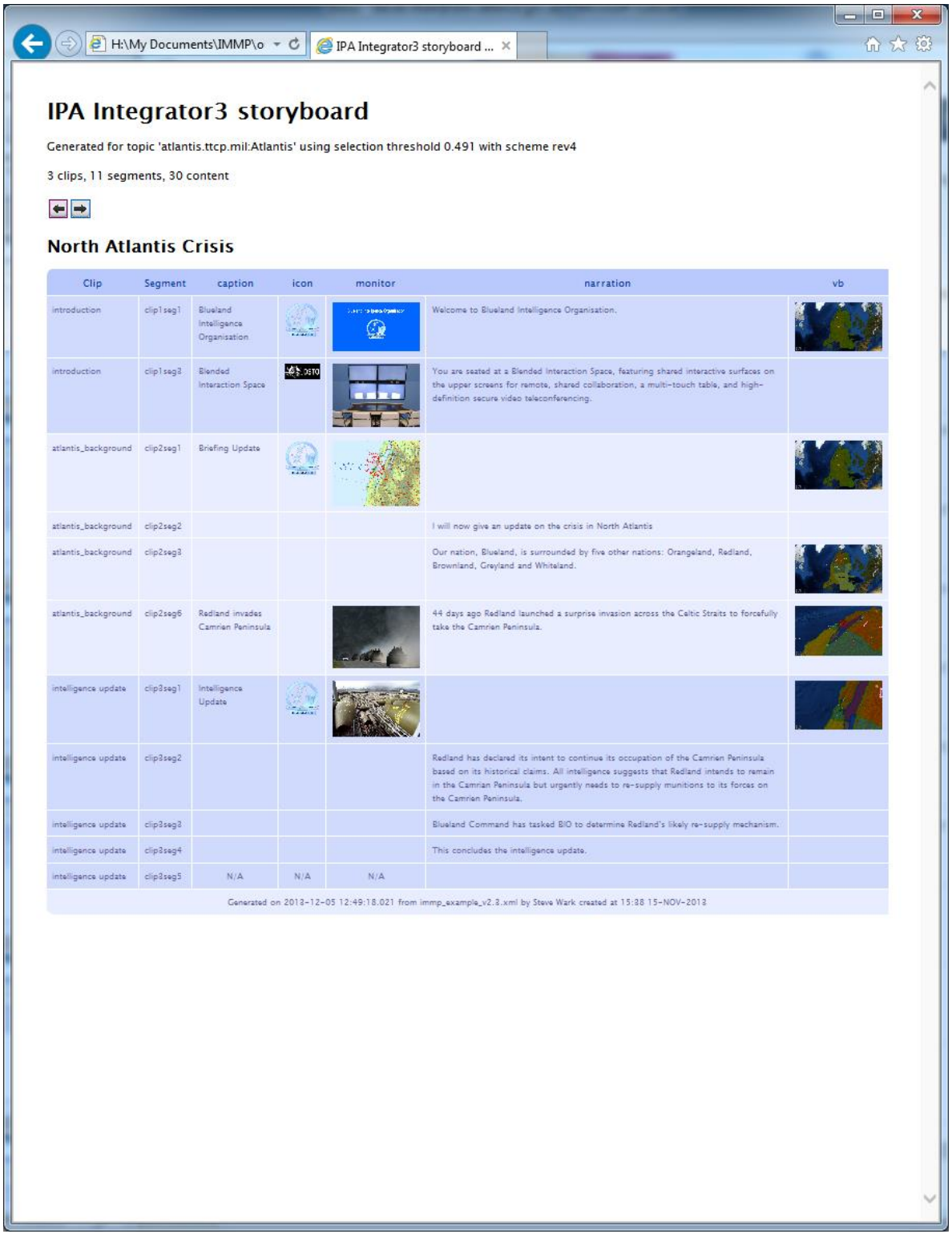
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.481 with scheme rev4

3 clips, 12 segments, 32 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg2	Virtual Adviser – Military			I am Jane, your Virtual Adviser on military content for today.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrian Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrian Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:17.959 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013



H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.491 with scheme rev4

3 clips, 11 segments, 30 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg3				Our nation, Blueland, is surrounded by five other nations: Orangeland, Redland, Brownland, Greyland and Whiteland.	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrien Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:18.021 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.578 with scheme rev4

3 clips, 10 segments, 28 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg6	Redland invades Camrien Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrien Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg2				Redland has declared its intent to continue its occupation of the Camrien Peninsula based on its historical claims. All intelligence suggests that Redland intends to remain in the Camrian Peninsula but urgently needs to re-supply munitions to its forces on the Camrien Peninsula.	
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:18.131 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.601 with scheme rev4

3 clips, 9 segments, 27 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	
intelligence update	clip3seg4				This concludes the intelligence update.	
intelligence update	clip3seg5	N/A	N/A	N/A		

Generated on 2013-12-05 12:49:18.209 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

H:\My Documents\IMMP\o IPA Integrator3 storyboard ...

IPA Integrator3 storyboard

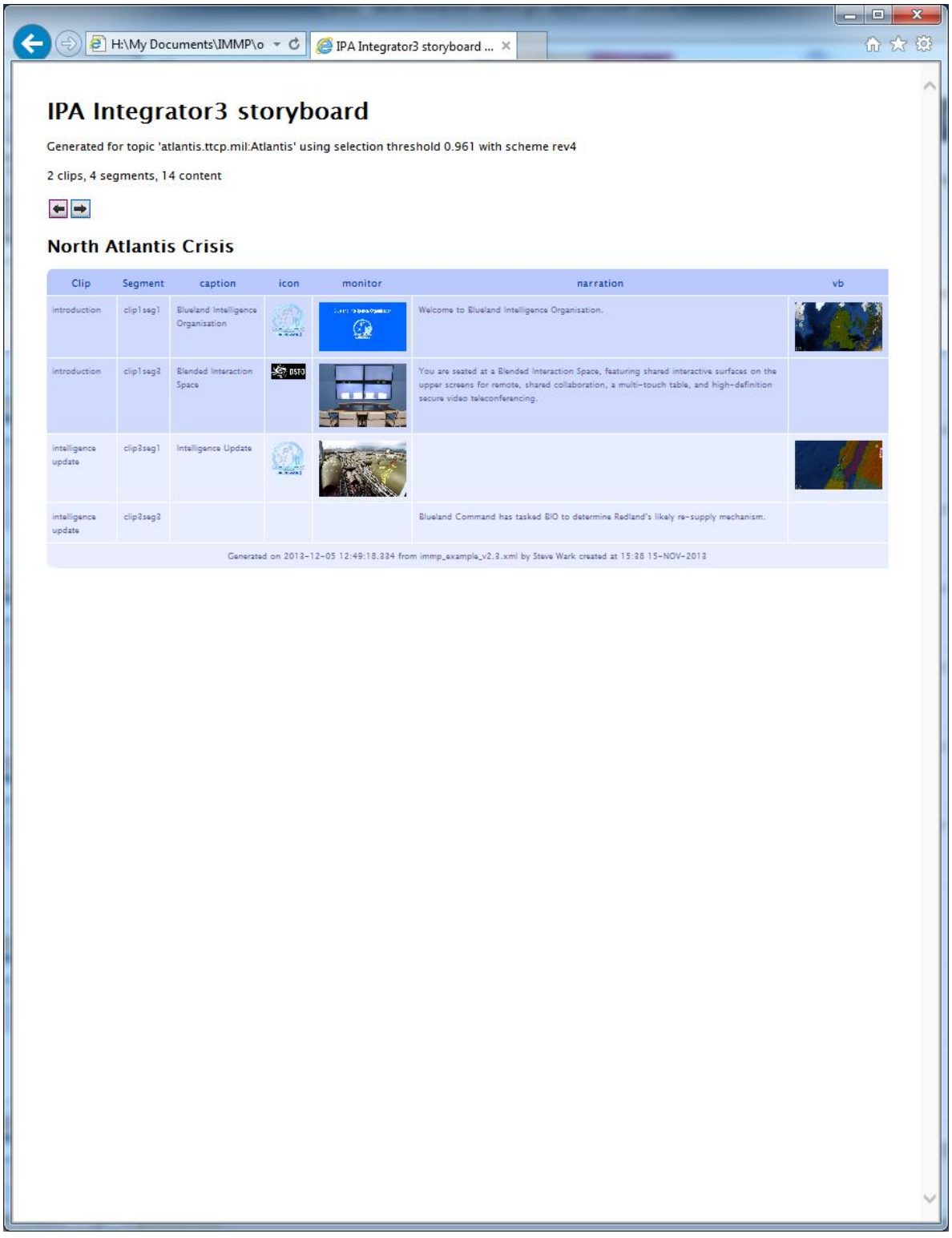
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.701 with scheme rev4

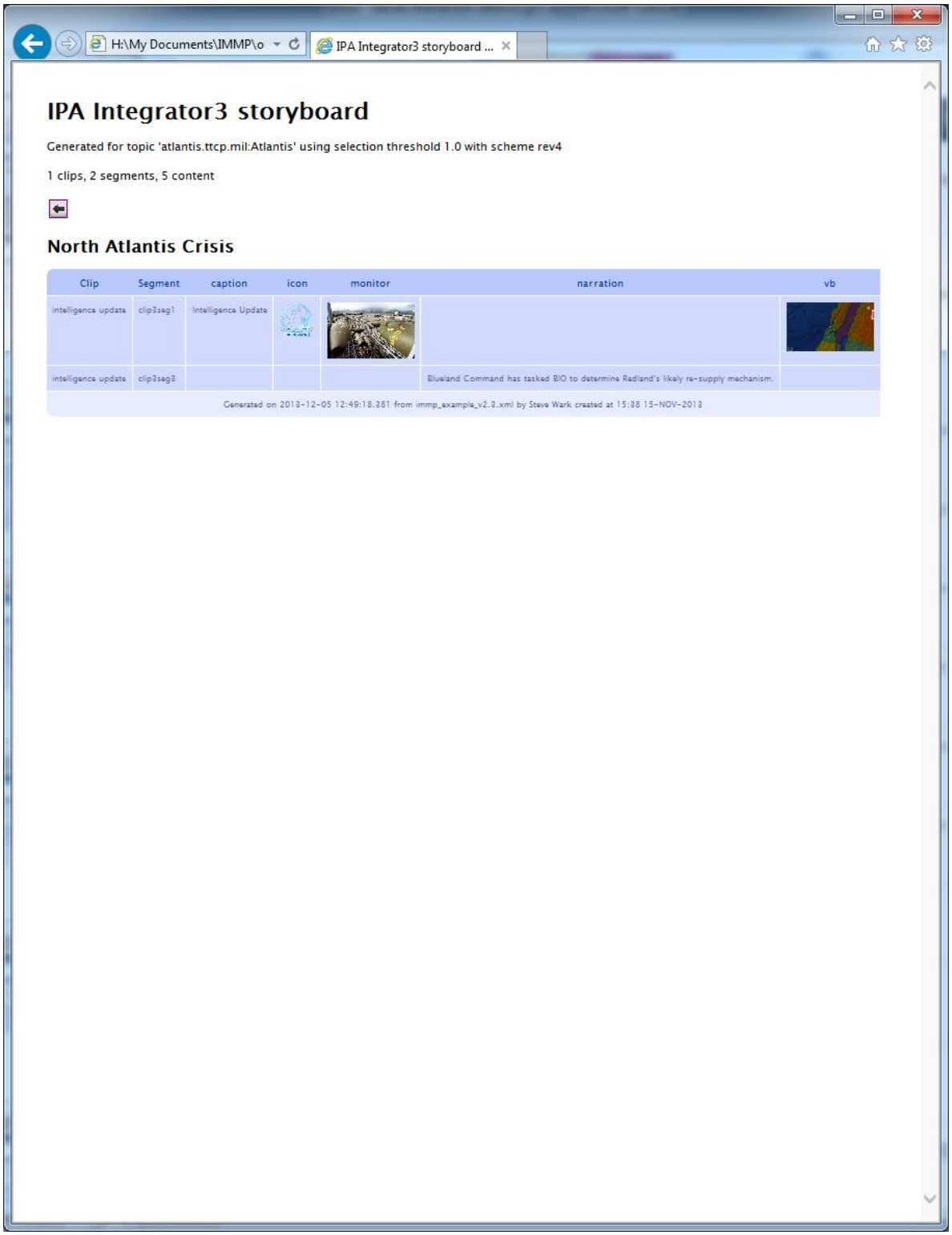
3 clips, 7 segments, 23 content

North Atlantis Crisis

Clip	Segment	caption	icon	monitor	narration	vb
introduction	clip1seg1	Blueland Intelligence Organisation			Welcome to Blueland Intelligence Organisation.	
introduction	clip1seg3	Blended Interaction Space			You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.	
atlantis_background	clip2seg1	Briefing Update				
atlantis_background	clip2seg2				I will now give an update on the crisis in North Atlantis	
atlantis_background	clip2seg6	Redland invades Camrian Peninsula			44 days ago Redland launched a surprise invasion across the Celtic Straits to forcefully take the Camrian Peninsula.	
intelligence update	clip3seg1	Intelligence Update				
intelligence update	clip3seg3				Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.	

Generated on 2013-12-05 12:49:18.287 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013


H:\My Documents\IMMP\o
IPA Integrator3 storyboard ...
IPA Integrator3 storyboard
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 0.961 with scheme rev4
2 clips, 4 segments, 14 content
North Atlantis Crisis
Clip
Segment
caption
icon
monitor
narration
vb
introduction
clip1seg1
Blueland Intelligence Organisation
Welcome to Blueland Intelligence Organisation.
introduction
clip1seg3
Blended Interaction Space
You are seated at a Blended Interaction Space, featuring shared interactive surfaces on the upper screens for remote, shared collaboration, a multi-touch table, and high-definition secure video teleconferencing.
intelligence update
clip3seg1
Intelligence Update
intelligence update
clip3seg3
Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.
Generated on 2013-12-05 12:49:18.334 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013


H:\My Documents\IMMP\o
IPA Integrator3 storyboard ...
IPA Integrator3 storyboard
Generated for topic 'atlantis.ttcp.mil:Atlantis' using selection threshold 1.0 with scheme rev4
1 clips, 2 segments, 5 content
North Atlantis Crisis
Clip
Segment
caption
icon
monitor
narration
vb
intelligence update
clip3seg1
Intelligence Update
intelligence update
clip3seg3
Blueland Command has tasked BIO to determine Redland's likely re-supply mechanism.
Generated on 2013-12-05 12:49:18.381 from immp_example_v2.3.xml by Steve Wark created at 15:38 15-NOV-2013

Appendix E: IMMP Prototype Quick Start Guide

E.1. Obtaining the software

The prototype IMMP software is available for download as a source bundle at the JOAD Decision Science software repository, as the *source.zip* artifact[29] of the *dsto.immp.scripts* project at:

http://c2-maven.dsto.defence.gov.au/nexus

When downloaded this should be a file of the form: *dsto.immp.scripts-XXX-sources.zip*

E.2. Installation

The prototype IMMP system can be installed on both Windows and Linux operating systems, but requires Java 7.

Once the latest *dsto.immp.scripts-XXX-sources.zip* artefact has been downloaded, to install it:

1. Unzip the file and install it into a suitable location on the file system of:
   - a client machine – on which you will be running the IMMP software
   - a VA service host – on which the Virtual Adviser service is installed, and which will be the machine rendering and displaying the presentation.

   These steps should have created a *dsto.immp.scripts* directory on these two machines, which contains the source scripts, dependencies, and default configuration files.
2. On the VA service host, set an environment variable 'IMMP_ROOT' to point to the directory in which the bundle was installed (*dsto.immp.scripts*). This is needed so that the appropriate configuration files can be loaded.

E.3. Testing

If you have gradle installed it is possible to test your installation of the prototype by running from the *dsto.immp.scripts* directory:

```
gradle test
```

This will test most of the functionality required for use of the IMMP prototype.

E.4. Configuration

[29] Latest information is available at: http://logwiki.dsto.defence.gov.au/display/va/IMMP

The way the VA renders the multimedia presentation can be changed by modifying the *config/immp.thml* file located in the *dsto.immp.scripts* directory on the VA service host. This allows adjustment of the VA character, scene layout, etc. This can also be modified by specifying a different initialisation file (again on the VA service host) using the *--initfile* command line parameter as described below. The default configuration for the VA includes the following channels:

- narration (for text)
- caption (for text)
- icon (for images to associate with caption)
- monitor (for images or video)
- title (for an automatically generated clip header)

For more information on the THML language used in the *config/immp.thml* file, refer to Appendix A.

E.5. Usage

To run the software use the 'immp.bat' file as per instructions below:

```
Usage: immp <file> [--topic=<ontology>:<topic>] [--out=<prefix>] [--style=<stylesheet>]
                   [--select=<threshold> | --duration=<secs>]
                   [--schema=<index> | --overview | --focussed]
                   [--storyboard | --graph[=<mode>] | --render[=<host>[:<port>]] |
                    --analyse]
                   [--initfile=<file> | --noinit] [--queue] [--sequence=<seqid>]
                   [--help] [--debug=<level>]
Where:
        <file> is XML file for IMMP presentation
        --help prints this help message
        --debug=<level> sets debug level, defaults to 0 (info)
        --topic=<ontology>:<topic> specifies ontology and topic to use for RST selection
        --out=<prefix> specifies filename prefix to use for output generation
        --style=<stylesheet> specifies name of stylesheet to use (not fully implemented)
        --select=<threshold> specifies selection threshold [0,1] to apply for IMMP content
        --duration=<secs> specifies upper limit on presentation duration
        --schema=<index> specifies selection schema to use based on index [1,4]
        --overview specifies selection schema retaining overall presentation structure
        --focussed specifies selection schema focussed on presentation nucleus
        --sequence=<seqid> selects the sequence with id <seqid>, defaults to all.
        --storyboard generates HTML storyboard for presentation
        --graph=<mode> generates a graphviz graph of the presentation using specified mode.
          Modes are 'symbolic', 'wsymbolic', 'preview', 'wpreview'. Defaults to 'preview'.
        --render[=<host>:<port>] generates and runs THML presentation on specified host
          Defaults to localhost:51627
        --initfile=<file> specifies initialisation file (on the rendering host) to use
        --noinit disables (re)loading of the initialization file (on the rendering host)
        --queue generates and sends the presentation as a queued presentation
        --analyse generates a summary report over the full range of selection thresholds
```

An example to generate and present a two minute example IMMP presentation, from the *dsto.immp.scripts* directory:

```
    immp src/test/resources/immp_example_v2.3b.xml "--duration=120" --render
```

To render the entire example presentation, use:

```
immp src/test/resources/immp_example_v2.3b.xml --render
```

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. DLM/CAVEAT (OF DOCUMENT)

2. TITLE

Intelligent Multi-Media Presentation Using Rhetorical Structure Theory

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document (U)
Title (U)
Abstract (U)

4. AUTHOR(S)

Steven Wark and Marcin Nowina-Krowicki

5. CORPORATE AUTHOR

DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend Victoria 3207 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-3067	AR-016-212	Technical Report	January 2015

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2014/1145289/1	CDF 07/362	DCJOPS	124	27

13. DSTO Publications Repository

http://dspace.dsto.defence.gov.au/dspace/

14. RELEASE AUTHORITY

Chief, Joint and Operations Analysis Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS

Situational Awareness, Intelligent Systems, Multimedia Systems, Natural Language Processing, Human Machine Interfaces

19. ABSTRACT

This report explores the feasibility of an Intelligent Multi-Media Presentation (IMMP) system for human-authored content, marked up using Rhetorical Structure Theory, to support dynamic selection of the presentation content based on user needs and preferences. It describes an XML format developed to represent an IMMP presentation, and a simplified prototype system developed to dynamically select and present content at different levels of detail within a specified maximum duration. An initial assessment of this system, based on the TTCP 'Military Strikes in Atlantis' scenario, found that it performed satisfactorily and that this is thus a feasible approach. Further work is planned to assess this system with other scenarios, and determine if it is a generally suitable approach.